# Chapter 7

# Upper Troposphere and Lower Stratosphere

**Lead Authors:** Andrew Gettelman & Michaela I. Hegglin

**Co-authors:** Seok-Woo Son
Masatomo Fujiwara
Simone Tilmes
Laura Pan
Peter Hoor
Huikyo Lee
Gloria L. Manney
Thomas Birner
Gabriele Stiller
Markus Rex
Stefanie Kremser
Don Wuebbles
Kaley A. Walker
Juan A. Añel

## 7.1 Introduction

The upper troposphere/lower stratosphere (UTLS) plays a key role in radiative forcing and chemistry-climate coupling (see Shepherd (2007) for a recent review). The UTLS is the region lying between the lower troposphere and the middle stratosphere, from roughly 5 to 22 km altitude. The dynamical, chemical, and radiative properties of the UTLS are in many ways distinct from both the lower troposphere and the middle stratosphere. The coupling between dynamics, chemistry, and radiation is especially strong in the UTLS, controlled by complex processes on a wide range of length and time scales. UTLS processes depend crucially on the distribution of greenhouse gases

(GHGs), especially $O_3$ and $H_2O$, as well as aerosols and clouds. Perturbations to the distributions of these atmospheric constituents can lead to direct forcing of surface climate through both radiative and dynamical mechanisms. The extra-tropical tropopause region is an important source of baroclinic instabilities that impact surface weather. In turn, climate change, through changing temperatures and transport patterns, has the potential to affect the chemical composition and structure of the UTLS.

In order to investigate the mechanisms determining the structure of the UTLS and to quantify future changes using CCMs, it is important that CCMs accurately represent the dynamical, radiative, and chemical properties of the UTLS. In this chapter, we present the first comprehensive validation of CCMs in the UTLS, using a wide range of process-oriented model diagnostics. To achieve this goal, new data sets are compiled and analysed in order to test their usefulness in serving as observational references. Many of the diagnostics are based on seasonal cycles, a long-established tool for validating models. Many other diagnostics presented here are used for the first time, and may require further development. The different diagnostics are used to grade model skill, which are summarized at the end of the chapter into a qualitative overall assessment of each model's performance. Some of the more complex diagnostics are applied only to a subset of the CCMVal-2 models that provided temporally higher-resolved instantaneous chemical and dynamical fields. These evaluations add additional information on key aspects of transport and dynamics in the UTLS, and may be regarded as examples of how future validation efforts could be expanded, but are not at this time comprehensive quantitative metrics of model performance. Some of the analyses are compared to recently published studies using the CCMVal-1 models. Also, past and future trends of key dynamical and chemical quantities are presented at the end of the chapter.

In many cases, a different balance of processes and structures exists in the tropics and at higher latitudes, providing a natural separation between the tropical UTLS, which also contains the Tropical Tropopause Layer (TTL), and the extra-tropical UTLS. We use this distinction as a natural break, but do not neglect the interactions across latitudes, and processes in the subtropics. **Table 7.1** provides an overview of the model diagnostics described in this chapter organised according to the key processes the diagnostics are testing for both the tropical and the extra-tropical UTLS.

The different diagnostics are explained in detail in the tropical and extra-tropical UTLS sections, but here a short summary is given:

**Tropical UTLS diagnostics:** For the tropical UTLS, or Tropical Tropopause Layer (TTL), we focus on several diagnostics of temperature, transport and water vapour. The TTL is the source of almost all stratospheric air, and water vapour in the stratosphere is regulated by tropopause temperatures (Brewer 1949). Tropopause temperature is an important aspect of model representation of the TTL since it has strong implications for the water vapour distribution. Other diagnostics focus on variability in the TTL, both for examining large scale and long term variability in tropopause temperature, as well as intra-seasonal variability and the representation of tropical wave modes in the models. Quantitative grades are reported for diagnostics for water vapour, tropopause temperature, and tropopause pressure.

**Extra-tropical UTLS diagnostics:** For the extra-tropical UTLS, we focus on several diagnostics of dynamics, transport, mixing and variability. The mass flux into the lowermost stratosphere (LMS, see **Figure 7.1**) from above and the seasonality in LMS mass determine the amount and temporal variability of stratospheric ozone transported into the troposphere, thereby having a crucial impact on the radiative budget of the upper troposphere, but also on tropospheric chemistry. The distributions of radiatively active species such as ozone and water vapour influence temperatures, winds, and dynamics in the extra-tropical UTLS. $O_3$ and $H_2O$ are key in determining the models' capabilities to represent stratosphere-troposphere coupling accurately. Several diagnostics therefore focus on how well the models represent the dynamical and chemical structure of the extra-tropical UTLS, especially the distributions of temperature, $O_3$ and $H_2O$.

The chapter starts with a description of the data sets used in the comparisons (Section 7.2), followed by an introduction to the diagnostics used in this chapter (Section 7.3). The main validation exercise is divided into two sections discussing UTLS characteristics of the tropics (Section 7.4) and the extra-tropics (Section 7.5) separately. In Section 7.6, we discuss past and future changes simulated in the models, before we summarize our findings and provide an overall assessment of the models' performance in the UTLS in Section 7.7.

## 7.2 Description of observational data sets used for CCM validation

High quality measurements in the global UTLS for the use of model validation are difficult to obtain due to major challenges for the available measurement platforms. *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

**Table 7.1:** List of core processes to validate CCMs in the UTLS. Gray highlights the diagnostics that will be used as quantitative metrics for the overall model assessment.

| Process | Diagnostic | Variables | Data | References[a] | Section |
|---|---|---|---|---|---|
| ***Tropical UTLS*** | | | | | |
| Dynamics | Seasonal cycle in CPT[b] | *T* | NCEP, ERA-40 | *Eyring et al. (2006)* | 7.4.1 |
| | TP inversion layer | *T* | GPS | *Gettelman et al. (2010)* | 7.4.7 |
| Dehydration | Seasonal cycle in $H_2O$ above CPT (80 hPa) | $H_2O$ | HALOE | *Eyring et al. (2006)* | 7.4.5 |
| | $H_2O$ – CPT correlations | $H_2O$ | NCEP, ERA-40, HALOE | *Gettelman et al. (2010)* | 7.4.5 |
| | Lagrangian CPT | *u, v, T*, Heating | NCEP, ERA-40 | *Kremser et al. (2009)* | 7.4.3 |
| Variability | Interannual CPT anomalies | *T* | NCEP, ERA-40 | *Gettelman et al. (2009)* | 7.4.1 |
| | Wave analyses | *T, u, v*, OLR[b] | ERA-40, ERA-Interim, NCEP, NCEP2, JRA25 | *Wheeler and Kiladis (1999)* | 7.4.6 |
| Transport & mixing | $O_3$ seasonal cycle (100 hPa) | $O_3$ | $O_3$-sondes | *Eyring et al. (2006)* | 7.4.4 |
| | Lagrangian Transport Time | *u, v, T*, Heating | NCEP, ERA-40 | *Kremser et al. (2009)* | 7.4.3 |
| ***Extra-tropical UTLS*** | | | | | |
| Dynamics | Zonal mean zonal wind @200 hPa | *u* | ERA-40, NCEP | *Hegglin et al. (2010)* | 7.5.1.1 |
| | Seasonal cycle in LMS mass | *M* | NCEP | *Appenzeller et al. (1996)* | 7.5.1.2 |
| | TP pressure anomalies | pressure | ERA-40, NCEP | *Gettelman et al. (2010)* | 7.5.1.3 |
| | TP inversion layer | *T* | GPS | *Birner (2006)* | 7.5.1.4 |
| Transport & mixing | Seasonal cycle in $O_3$, $HNO_3$, $H_2O$ @100 and 200 hPa | $O_3$, $HNO_3$, $H_2O$ | MIPAS, ACE-FTS, MLS | *Logan (1999)* | 7.5.2.1 |
| | Meridional tracer gradients @200 hPa | $O_3$ | MLS | *Shepherd (2002)* | 7.5.2.2 |
| | Normalised CO relative to TP | CO | SPURT | *Hoor et al. (2004, 2005); Hegglin et al. (2010)* | 7.5.2.3 |
| | Vertical profiles in TP coordinates | $H_2O$, CO, $O_3$ | Aircraft, ACE-FTS | *Pan et al. (2004, 2007) Tilmes et al. (2010)* | 7.5.2.4 |
| | ExTL[b] depth | $H_2O/O_3$ | Aircraft, ACE-FTS | *Pan et al. (2007) Hegglin et al. (2009)* | 7.5.2.5 |
| Variability | PDFs of $O_3$ variability | $O_3$ | MLS | *Rood et al. (2000)* | 7.5.3 |

[a] *Listed references provide information on the diagnostic and/or the observations used for the evaluation*

[b] *Abbreviations: CPT=cold point temperature; TP=tropopause; OLR=outgoing long-wave radiation; ExTL=extra-tropical tropopause transition layer*

length scales found in the chemical and dynamical fields in the UTLS – the result of the large dynamical activity 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. Note that the list below is not a comprehensive compilation of currently available data sets which may be useful for model evaluation in the UTLS.

### 7.2.1 Balloon data

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 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 analysed (Seidel and Randel, 2006).

### 7.2.2 Aircraft data

As with balloon measurements, aircraft observations provide mostly high accuracy, high precision, and high resolution data in the UTLS, but may be restricted in their representativeness due to limited sampling in time and space.

Data from various NASA, NSF, and some German aircraft campaigns between 1995 and 2008 have recently been compiled into a high resolution aircraft based UTLS climatology of ozone, CO and $H_2O$ (Tilmes *et al*., 2010). The data set covers a broad altitude range up to 22 km. The spatial coverage ranges over all latitudes in the NH for most of the four seasons, but coverage is predominantly over North America and Europe. The precision and accuracy of the ozone data is ± 5%. CO observations taken by different instruments have a precision of < 1% and an accuracy of < 3%. The precision of $H_2O$ data is estimated to be < 5% and the accuracy is between 0.3 ppmv and values of 10% depending on the instrument. The aircraft climatology is especially designed to serve as a tool to evaluate the representation of chemistry and transport by CCMs in the UTLS.

A subset of these high-resolution and high-precision observations is used separately in this chapter and stem from the German SPURT aircraft campaign (SPURenstofftransport in der Tropopausenregion, or trace gas transport in the tropopause region). The campaign consisted of 8 deployments distributed seasonally over the course of three years (2001-2003), with a total of 36 flights, each yielding around 2-5 hours of observations. The flights were carried out between around 35°N and 75°N over Europe and reached potential temperature levels between 370 K and 375 K. A campaign overview is given by Engel *et al*. (2006). The CO measurements used in this study typically showed total uncertainties of 1.5% (Hoor *et al*., 2004).

Another subset of high-resolution data used separately in this chapter stems from the NASA POLARIS (Photochemistry of Ozone Loss in the Arctic Region in Summer) campaign (Newman *et al*., 1999). During the campaign, 35 flights were deployed between March and September 1997 using the NASA ER-2 research aircraft from three locations: Moffett Field, California (~37°N), Fairbanks, Alaska (~65°N), and Barbers Point, Hawaii (~21°N). The flights covered a latitude range of approximately 20°N-70°N and a vertical range of 5-18 km. For the $O_3$ and $H_2O$ data used in this study, the estimated accuracies are ~3% and 5%, respectively (Proffitt and McLaughlin, 1983; Hintsa *et al*., 1999). The use of these data to characterize the ExTL has been described in Pan *et al*. (2004; 2007).

### 7.2.3 Satellite data

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. To determine the accuracy and precision of these measurements is the focus of intensive validation efforts. While more data sets will become available in the near future, here we describe only the data from instruments used in this chapter.

#### *ACE-FTS on SCISAT-1*

The Atmospheric Chemistry Experiment Fourier Transform Spectrometer (ACE-FTS) on Canada's SCISAT-1 satellite features high resolution (0.02 $cm^{-1}$) and broad spectral coverage in the infrared (750 to 4400 $cm^{-1}$) (Boone *et al*., 2005; Bernath *et al*., 2005). The instrument has operated since February 2004 in solar occultation mode providing seasonally varying coverage of the globe, with an emphasis on mid-latitudes and the polar regions. Up to 30 occultation events occur per calendar day. The very high signal-to-noise ratio characterizing the ACE-FTS infrared spectra makes it possible to measure more than 30 chemical trace gas species with high accuracy and precision (Clerbaux *et al*., 2008; Dupuy *et al*., 2008; Hegglin *et al*., 2008). The derived overall measurement uncertainties in the observations for the UT and LS were ±9% and ±12% for CO, ±30% and ±18% for $H_2O$, and ±18% and ±8% for $O_3$, respectively (Hegglin *et al*., 2008). This, together with vertical sampling ranging from about 3 km to less than 1 km in the UTLS, provides the first global view of tracer distributions in the extra-tropical tropopause region (Hegglin *et al*., 2009).

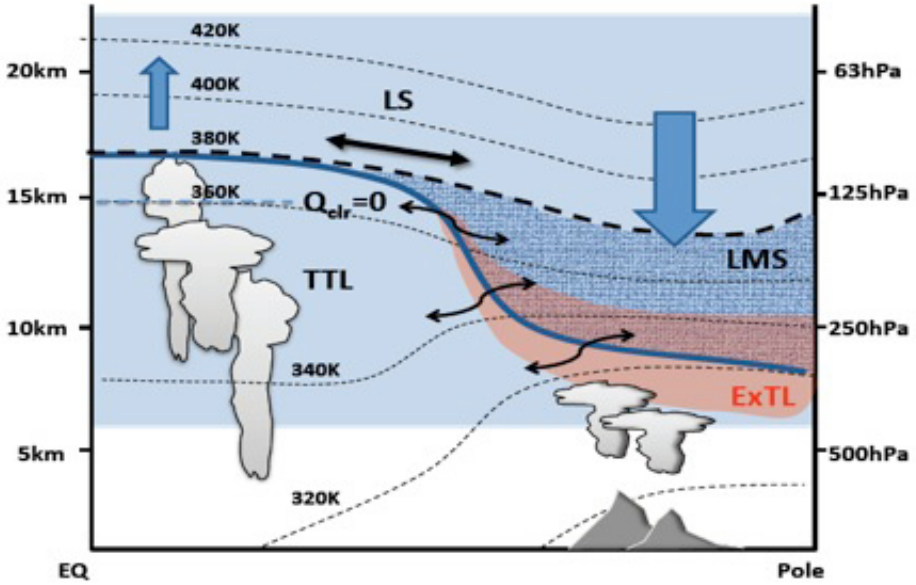


**Figure 7.1:** Schematic of the UTLS illustrating the Tropical Tropopause Layer (TTL) extending from the level of deep convective outflow, through the level of zero clear-sky radiative heating ($Q_{clr}$=0) up to the cold point and the tropopause (thick blue line). The lower stratospheric (LS) branch of the Brewer-Dobson wave driven circulation is shown by the blue arrows. Faster and seasonally dependent meridional transport between the tropics and extra-tropics is found within the 380-420 K layer (black straight arrow). The extra-tropical UTLS contains the lowermost stratosphere (LMS) between the 380 K potential temperature surface and the tropopause. Also illustrated (in red) is the Extra-tropical Transition Layer (ExTL) which represents a mixing layer in which air has partly tropospheric, partly stratospheric chemical characteristics.

*Aura MLS*

The Microwave Limb Sounder (MLS) on the EOS Aura satellite measures millimeter- and submillimeter-wavelength thermal emission from the limb of Earth's atmosphere (Waters *et al*., 2006). Aura MLS has data coverage from 82°S to 82°N latitude on every orbit, providing comprehensive information on UTLS tracer distributions. Vertical profiles are measured every 165 km along the sub-orbital track and have a horizontal resolution of ~200–300 km along-track and ~3–9 km across-track. Vertical resolution of the Aura MLS data is typically ~3–4 km in the lower and middle stratosphere (Livesey *et al*., 2007). $O_3$ has been used successfully in studies to examine transport in the UTLS, although some biases still exist in the version 2.2 retrieval used in the evaluations presented here. Validation of stratospheric $O_3$ is discussed by Livesey *et al*. (2008). The MLS v2.2 $O_3$ retrieval has an accuracy of 0.02 ppmv and 0.05 ppmv at 214 hPa and 100 hPa, respectively, and a precision of 0.04 ppmv at both levels.

*MIPAS*

MIPAS is a limb-viewing Fourier transform emission spectrometer on board Envisat in a sun-synchronous polar orbit. MIPAS covers the mid-infrared spectral region between 685 and 2410 $cm^{-1}$ (Fischer *et al*., 2008). MIPAS has provided data since 2002 at about 1000 geo-locations per day from pole to pole during day and night. MIPAS covers the atmosphere from the upper troposphere to the mesosphere (6 to 70 km), and provides global distributions of a large number of species. In its original observation set-up from July 2002 to March 2004 it measured one limb radiance profile every 500 km along track with a vertical sampling of 3 km and a spectral resolution of 0.035 $cm^{-1}$. Validation of these data products can be found in Milz *et al*. (2005, 2009), Wang *et al*. (2007), and Steck *et al*. (2007). Since January 2005, the observation set-up has been changed to slightly reduced spectral resolution (0.0625 cm-1), but improved vertical (1.5 km) and horizontal along-track (400 km) sampling. Description of these data products can be found in von Clarmann *et al*. (2009). The accuracy (including contributions of precision and systematic errors) of the MIPAS data has been found to be 9.6% (at 15 km) and 17% (at 10 km) for $O_3$, 4.4% (at 15 km) and 6.0% (at 10 km) for $HNO_3$, and 17.7% (at 20 km) and 8.3% (at 15 km) for $H_2O$, respectively (von Clarmann *et al*., 2009). All data used within this study have been processed at the Institute for Meteorology and Climate Research (IMK) (von Clarmann *et al*., 2003).

*HALOE*

We also use water vapour 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.,* Kley *et al*., 2000). HALOE validation and a 13-year record (1992-2004) gives us high confidence in HALOE performance. The random and systematic errors in HALOE $H_2O$ at 100 hPa are 11% and 28%, and for $O_3$ 14% and 24%, respectively.

*COSMIC*

The Global Positioning System (GPS) Radio Occultation (RO) data used in this study were obtained from the COSMIC (Constellation Observing System for Meteorology, Ionosphere, and Climate)/FORMOSAT-3 (Formosa Satellite Mission 3) mission, which is a collaborative project between Taiwan and the United States (Anthes *et al*., 2008). The mission placed six micro-satellites in different orbits at 700-800 km above the ground. These satellites form a low-orbit constellation that receives signals from US GPS satellites, providing approximately 2500-3000 soundings per day almost evenly distributed over the globe. The mission has a relatively short data record since its mission launch was only in 2006. In this study, we use data between 2006 and 2009.

### 7.2.4 Meteorological Analyses

Operational meteorological analyses are produced on a daily basis by weather forecast centres. These analyses (or 'reanalyses' 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 Center 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 (JRA25) described by Onogi *et al*. (2007), the European Centre for Medium Range Weather Forecasts (ECMWF) 40 year reanalysis (ERA-40) described by Uppala *et al*. (2005) and 'interim' analysis (ERA-Interim) described by Uppala (2008). For information on the different reanalyses (ERA-40, NCEP, JRA25) the reader is referred to Randel *et al*. (2002) and references therein. A few distinct caveats common to reanalyses 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 reanalysis 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, reanalysis 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 (Pawson and Fiorino, 1998). Thus, the reanalyses need to be treated with some caution (Randel *et al*., 2002). For comparison purposes with temperature and the tropopause, we will use the ERA-40 reanalysis, because of their high quality and a relatively long (20 year) record for comparison.

## 7.3 Metrics and Grading

Metrics are diagnostics with quantitative grading (a 'grade') applied, and are used to quantitatively assess model behaviour for some of the diagnostics. For example, mean values of a certain quantity or the amplitude and phase of a seasonal cycle can be used as a metric. We introduce below the two main approaches that were used in this chapter to lead from those metrics to a grading of the models. If a different grading approach is used for a diagnostic, it will be explained in the respective sections. Section 7.7 discusses how these approaches are applied and what values are 'acceptable'.

### 7.3.1 Grading of Mean and Correlative Quantities

Some metrics (*e.g*., in the tropics for the cold point tropopause temperature, water vapor annual cycle and tropopause pressure, and in the extra-tropics for the zonal mean wind) are defined following Douglass *et al*. (1999) and Waugh and Eyring (2008), with extensions to look at variability. Metrics are based on defining monthly means after spatial averaging. Douglass *et al*. (1999) define a metric based on mean differences:

$$g_m = max(0, 1 - \frac{1}{n}\sum_{i=0}^{n}\frac{|\mu_{obs} - \mu_{mod}|}{n_g \sigma_{obs}}). \quad (7.1)$$

Here, $\mu$ is a monthly mean quantity and $n_g$ a scaling factor representing a number of standard deviations ($\sigma$), often taken to be 3 (Waugh and Eyring, 2008). We also define a metric based on correlated variability where $\mu'$ are anomalies from a mean quantity and $C$ is the linear correlation coefficient:

$$g_c = (Cor(\mu'_{mod}, \mu'_{obs}) + 1)/2. \quad (7.2)$$

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

We can also define a metric based on the magnitude of the variance of a quantity, where $\sigma$ is calculated each

month:

$$g_v = max(0, 1 - \frac{1}{n}\sum_{i=0}^{n}\frac{|\sigma_{obs} - \sigma_{mod}|}{n_g \sigma_{obs}}). \quad (7.3)$$

A single metric is then the linear combination:

$$G_{mod} = (g_m + g_c + g_v)/3. \quad (7.4)$$

This was found to yield reasonable results and encapsulate more than just the difference in means. 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 metrics recently identified by Grewe and Sausen (2009).

We have evaluated grades using several different measures of $\sigma_{obs}$ and $\mu_{obs}$ from different reanalysis systems, or using $\sigma_{obs}$ and $\mu_{obs}$ estimated from an ensemble of reanalysis 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 for some idea of the potential spread. We also examine the multi-model mean, calculated by summing model outputs to generate a multi-model $\mu_{obs}$. The multi-model grades and averages are also shown. The goal of applying grades is to quantitatively determine model deficiencies with sufficient detail to understand where models perform well and why models do not perform well.

### 7.3.2 Taylor Diagram

Taylor diagrams (Taylor, 2001) are used as an alternative to the above metrics in the extra-tropics and visualise the statistical summary of how well two patterns from a test field ($f$) and a reference field ($r$) match each other in terms of their correlation ($R$), their root-mean-square (RMS) difference ($E'$), and the ratio of their variances ($\sigma_f / \sigma_r$). These quantities can be used to quantify the correctness of phase and amplitude of seasonal cycles, which are often used in model-measurement comparisons.

The ratio between the variances of the test and reference field $\sigma_f / \sigma_r$ is the normalised variance of the test field and given as the radial coordinate in the Taylor diagram. It therefore corresponds to the distance between $f$ and the origin of the plot (see **Figure 7.2**).

The correlation $R$ is defined by

$$R = \frac{\frac{1}{N}\sum_{n=1}^{N}(f_n - \bar{f})(r_n - \bar{r})}{\sigma_f \sigma_r}, \quad (7.5)$$

and is given as the azimuthal coordinate in the Taylor diagram. In principle, $R$ could be negative and represented in a Taylor diagram (see Taylor, 2001 for an example). The RMS difference is defined by

$$E' = \sqrt{\frac{1}{N}\sum_{n=1}^{N}[(f_n - \bar{f}) - (r_n - \bar{r})]^2} \quad (7.6)$$

and can be determined once the correlation $R$ and $\sigma_f / \sigma_r$ are known. Smaller $E'$ represents a better fit between the test field and the reference field. Taylor diagrams can be used to test various aspects of model performance and to specify the relative skill of many different models (Taylor, 2001) as for example used in the Houghton *et al*. (2001) or the chemistry transport model inter-comparison by Brunner *et al*. (2003; 2005). The skill factor ($S$) can be defined using

$$S = \frac{4(1+R)}{(\hat{\sigma}_f + 1/\hat{\sigma}_f)^2(1+R_0)}. \quad (7.7)$$

Here, $\hat{\sigma}_f$ is the standard deviation of the test field normalised by the standard deviation of the reference field ($\sigma_f/\sigma_r$). $R_0$ is the maximum correlation models can achieve. Choosing $R_0 < 1$ allows us to account for uncertainty in the observations or model limitations such as spatial and temporal resolution. The skill approaches unity as the model

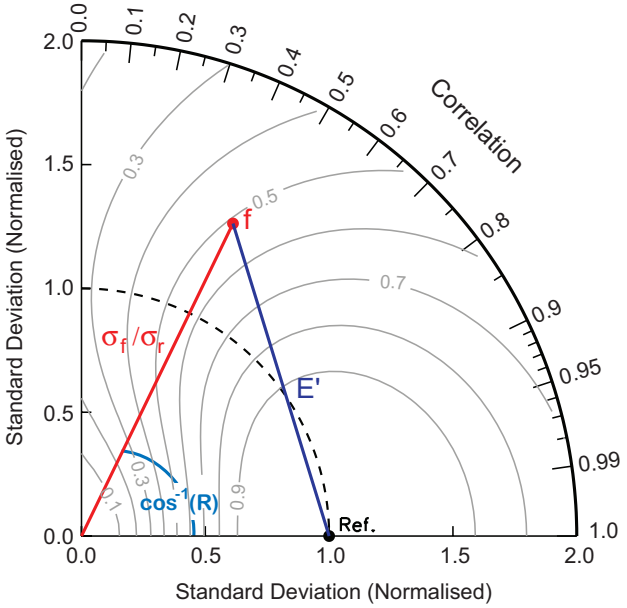


**Figure 7.2:** A sample Taylor diagram. f indicates the test or model field, and Ref. (or r) a reference field. The inverse cosine of the correlation R between the test and reference field (indicated in light blue) determines the location on the azimuthal axis. The radial distance of f from the origin corresponds to the standard deviation of the test field normalised by the standard deviation of the reference field ($\sigma_f / \sigma_r$, red line). The RMS difference (E', dark blue line) between test and reference field is proportional to the distance between the two fields on the diagram. Grey thin lines indicate the skill score (S) of the test field, which obtains in this example a value of 0.51.

variance approaches the observed variance (*i.e.*, as $\sigma_f \rightarrow 1$) and as $R \rightarrow R_0$.

For another illustrative example of how to read a Taylor diagram see also Hegglin *et al.* (2010). Finally, Taylor diagrams do not yield information on how close the mean of a given test field is to that of the reference field. Equation (7.1) is therefore used in addition to the skill to grade the simulated mean value ($g_m$). A single grade for a model is then the linear combination of the skill and the mean:

$$G_{tot} = (S + g_m) / 2. \quad (7.8)$$

## 7.4 Results: The Tropical UTLS

The tropical UTLS region is usually known as the Tropical Tropopause Layer (TTL). The TTL is the region in the tropics within which air has characteristics of both the troposphere and the stratosphere. The tropical tropopause layer sets the lower boundary condition for the stratosphere (Brewer, 1949). Representing the TTL correctly in global models is critical for simulating the future of the TTL and its effects on climate and chemistry of the stratosphere.

The TTL is the layer in the tropics between the level of main convective outflow and the cold point, about 12–19 km (Gettelman and Forster 2002). The TTL has also been defined as a shallower layer between 15–19 km (Fueglistaler *et al.*, 2009). We will use the deeper definition of the TTL here because we seek to understand not just the stratosphere, but the tropospheric processes that contribute to TTL structure (see below), and it is more representative of the tropical UTLS. 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 chemical boundary conditions of the stratosphere are set in the TTL. Clouds in the TTL, both thin cirrus clouds and convective anvils, have a significant impact on the radiation balance (Gettelman *et al.*, 2004; Corti *et al.*, 2006).

Here we will explore the representation of the TTL in global models, and assess potential changes to the TTL over time in Section 7.6.2.

### *Detailed Description of Tropical Diagnostics*

The following diagnostics also have quantitative metrics defined as noted in Section 7.3. We set $n_g = 3$ ($3\sigma$ threshold) for temperature and water vapour, and set the grading threshold ($3\sigma n_g$) to 10 hPa for tropopause pressure, because it represents 1 CCMVal-2 level around the tropopause.

- ***Diagnostic 1:*** Temperature of the Cold Point Tropopause (Amplitude and Phase of Annual Cycle): It is critical that models reproduce the amplitude and phase of the annual cycle of Temperature at the Cold Point Tropopause (TCPT), the coldest point in a UTLS profile. Because of the non-linearity of the Clausius-Clapeyron equation regulating water vapour saturation, the annual cycle is as important as the annual mean. This is a simplified metric of the true 'Lagrangian Cold Point', which we can examine in only a few models.
- ***Diagnostic 2:*** Tropopause Pressure: Lapse rate tropopause pressure (PTP) and its interannual variability should reflect the interannual variability in the observations. In particular, responses to major forcing events (*e.g.*, ENSO and volcanoes) should resemble observations. Anomalies of PTP have been shown to be more robust (*i.e.*, there is better agreement between observations) than tropopause temperature (Gettelman *et al.*, 2009). Simulated anomalies can be compared to reanalysis and radiosonde observations. A metric for this diagnostic is the correlation with interannual anomalies and the mean values from reanalysis systems in similar coordinates. A measure of the uncertainty is the variability of the grade of reanalysis systems compared to each other, which gives a sense of the unforced variation between analysis models.
- ***Diagnostic 3:*** Water vapour above the CPT (80 hPa): In conjunction with the CPT, the water vapour concentration above the CPT is the dominant term in the total hydrogen budget of the stratosphere. Models should simulate appropriately the annual cycle of the water vapour concentration in the lower tropical stratosphere, and its interannual variability.
- ***Diagnostic 4:*** Ozone in the TTL: Ozone in the TTL is affected by both transport and chemistry. TTL ozone is an important indicator of TTL processes. Models should represent the vertical structure of ozone and its annual cycle. The ozone concentration at a fixed level in the TTL (100 hPa) is used as a proxy for the ozone gradient. Ozone is also radiatively important in the TTL, and thus important for getting the thermal structure correct. 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 $NO_x$ production by lightning), which may or may not be present in a given model, but this needs to be understood.

### *Diagnostics not used as quantitative metrics in the overall model assessment*

These diagnostics are not used as quantitative metrics because they either do not yet have clear and robust

quantitative relevance that fits the grading methodology (Diagnostics 5 and 8), or are not produced for a large fraction of the models (Diagnostics 6 and 7).

- ***Diagnostic 5:*** $H_2O$ Correlations with TCPT: $H_2O$ at 80 hPa and TCPT can be combined by translating TCPT into water vapour by a saturation vapour mixing ratio. There should be a correlation between $H_2O$ and TCPT. This can also be expressed as the saturation vapour mixing ratio of the CPT ($Q_{SAT}(T_{CPT})$) and the ratio $H_2O$ /$Q_{SAT}(T_{CPT})$ should reflect physical mixing processes (*e.g.*, $H_2O$ /$Q_{SAT}(T_{CPT}) > 1$ is physically implausible based on observations). Monthly anomalies with a 1 month lag or interannual anomalies can also be compared.
- ***Diagnostic 6:*** Tropical Waves: The structures seen in transport times, chemistry and water vapour are not just consequences of large-scale processes on zonal or monthly scales. Waves and intra-seasonal variability in the TTL are critical for properly representing structures in the TTL. Here we use detailed high-resolution information to examine intra-seasonal variability and wave modes in a subset of models that provided high temporal frequency output (every 6 hours). Wave activity metrics are defined, but only a few models are analysed.
- ***Diagnostic 7:*** TTL transport time: The transport time through the TTL is a complex, emergent diagnostic reflecting a mix of transport processes, including large-scale advection by radiation and waves, as well as rapid convective motion in the vertical. Representing the transport time through the TTL is critically important 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 with high temporal resolution output to drive a trajectory model, and compare them to similar calculations with a reanalysis system.
- ***Diagnostic 8:*** Tropopause Inversion Layer: The Tropopause Inversion Layer (TIL) is a layer exhibiting an increase in the 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. Here we analyse the TIL in models that provided 3D instantaneous output.

### 7.4.1. Cold Point Tropopause Temperature

The TCPT is analysed from models and reanalyses vertically interpolated to CCMVal-2 levels. This provides a slightly blurred picture of the true cold point relevant for $H_2O$ condensation experienced by a simulated air parcel, but it is a useful baseline for comparisons. The annual cycle of tropical cold point temperature (or cold point tropopause temperature) is illustrated in **Figure 7.3** using the REF-B1 CCMVal-2 model fields. In addition to the models, several analysis systems are also shown (ERA-40, NCEP, NCEP2, JRA25, ERA-Interim). All analyses use monthly means interpolated to CCMVal-2 standard levels, so the models and analyses systems are on the same temporal and vertical grids. The gray region is 3 standard deviations ($\sigma$) from the ERA-40 observations. In general, almost all models are able to reproduce the annual cycle. There are significant differences between the models, but the monthly averages of 8 models and the multi-model mean are clustered within $3\sigma$ of the mean of ERA-40, as seen in the quantitative grades ($g_m$) in Figure 7.3. The quantitative metrics of the cold point are based on Equation (7.4). ERA-40 is taken as the base observation for the mean ($\mu_{obs}$) and the standard deviation ($\sigma_{obs}$). The reanalyses themselves do not all compare well (*i.e.*, score highly) compared to ERA-40, largely due to the warm bias of NCEP, and lack of correlated interannual variability.

The multi-model mean is very close to ERA-40, closer than some other analysis systems. These results are also better than CCMVal-1 models reported by Gettelman *et al*. (2009). Note that there is general quantitative agreement between the reanalyses, with 'grades' (compared to ERA-40) ranging from 0.6-0.8 (Figure 7.3). This is largely due to mean offsets between the analysis systems, the amplitude and phase of the annual cycle are in good agreement. NCEP and NCEP2 have a known warm bias (Pawson and Fiorino, 1998).

Most models do not show strong long-term trends in cold point temperatures, as indicated in **Figure 7.4**. NCEP and NCEP-2 reanalyses show strong cooling, which is not seen in the ERA-40, JRA25 and ERA-Interim analyses, as noted by Zhou *et al*. (2001). Note that these trends differ from other cooling trends reported from radiosondes (Gettelman and Forster, 2002; Seidel and Randel, 2006). This may be due to the gridding and interpolation to a standard set of vertical levels. Thus, if there is cooling of the cold point, it is not clear that this appears significantly in coarse resolution analyses. However, the lack of agreement among observations highlights the uncertainty in interannual and long-term changes in the TCPT.

Interannual variability is also illustrated in **Figure**

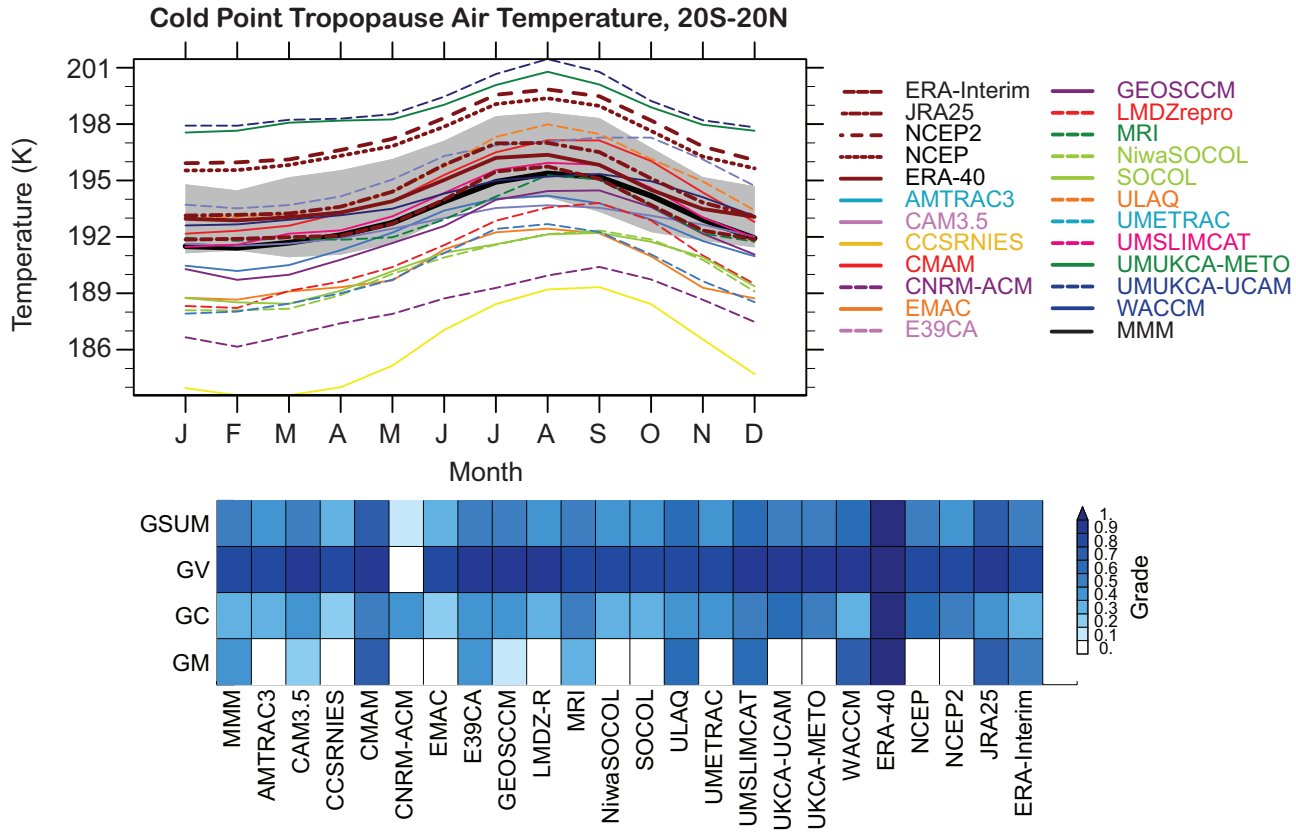


**Figure 7.3:** (Above) Annual cycle of tropical (20°S-20°N) cold point tropopause temperature from models and observations. Output and observations are from the period 1980-1999. Gray shaded region is 3σ variability from ERA-40 analyses. Multi-model mean is the thick black line, JRA25 (brown dot-dash), NCEP (brown dotted), NCEP2 (brown dashed), ERA-40 (brown solid), ERA-Interim (brown short-dash). (Below): Quantitative metrics summary of Cold Point Tropopause Temperature (TCTP) for mean (GM), correlation (GC), variance (GV) and the average of all three grades (GSUM).

**7.4**. Most models show significant warming of the cold point in 1991 of a degree or so, likely associated with the eruption of Mt. Pinatubo. Most models and the multi-model mean show warming of the TCPT after volcanic events (especially El Chichón in 1983 and Mt. Pinatubo in 1991). This warming is not as clear in analysis systems. Some models have a warming that is much too large (CNRM-ACM, NiwaSOCOL, SOCOL, MRI). This is factored into the metrics for variability ($g_v$) as described above. Models with large variability will have a lower score on the $g_v$ metric component (Figure 7.3). Also, the effect of the El-Niño Southern Oscillation (ENSO) and the Quasi-Biennial Oscillation (QBO) on the tropical tropopause (Zhou *et al*., 2001) are not clear in these low vertical resolution analyses. Interannual anomalies are not correlated between models and analyses, or between analyses themselves.

## 7.4.2 Lapse Rate Tropopause Pressure

The pressure of the lapse rate tropopause (PTP) has been shown to be a more robust metric than the cold point temperature (Gettelman *et al*., 2009). PTP is more sensitive to increasing thickness (vertically integrated temperature) below, and the temperature response is a more vertically confined. It is easier to get bulk thickness (latent heat release) right than TCPT details. This is evidenced by a high (0.8-0.9 or 1) correlation ($g_c$) among most analysis systems compared to ERA-40 (Figure 7.6). Grades are determined based on Equation (7.4). The meridional structure of tropopause pressure from models and analysis systems is shown in **Figure 7.5**. 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 at the 90 hPa level (when interpolated to CCMVal levels). Several models shift the tropopause up or down by a level. There are large differences, however, in the diagnosed tropopause at high latitudes, again due to potential shifts by a level or so in the thermal structure. CCMVal-2 levels are noted on Figure 7.5

Long-term changes in the tropopause pressure from 20°S-20°N are shown in **Figure 7.6**. There is good agreement between the interannual anomalies of most of the

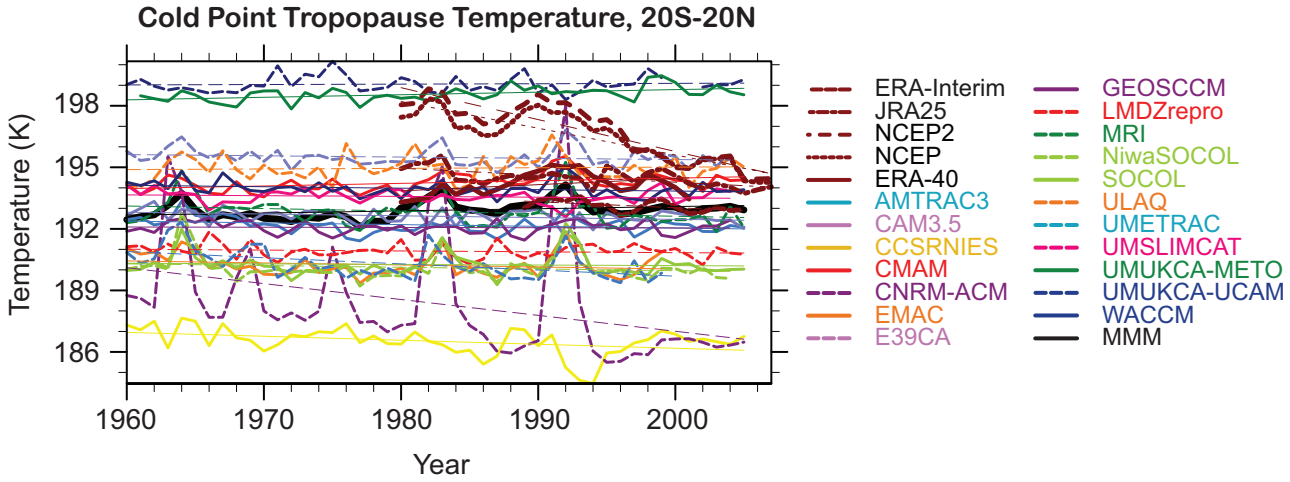


**Figure 7.4:** Time series of annual mean temperature of the Cold Point Tropopause (TCPT) for 20°S-20°N from models and analyses for 1960-2007. Thin lines are linear fits. Multi-model mean (MMM) is the thick black line.

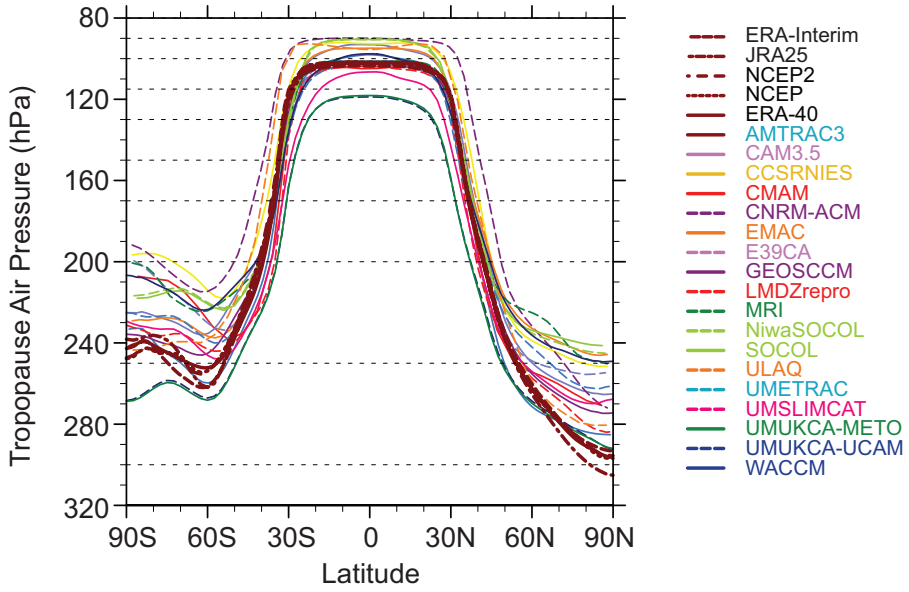


**Figure 7.5:** REF-B1 lapse rate tropopause pressure (PTP) annual zonal mean for 1980--1999 from models and analysis systems. Dotted lines represent CCMVal-2 vertical level structure in the UTLS, with levels at 400, 300, 250, 200, 170, 150, 130, 115, 100, 90, 80, 70, 50 hPa.

models, as well as trends in tropopause pressure. The simulated variability in models is higher than the observations. Most models and analysis systems show decreases in tropopause pressure associated with volcanic events, though model variability is larger. In particular, it is too large for CNRM-ACM, which jumps 2 levels (90 to 115hPa). Metrics for tropopause pressure 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.

### 7.4.3 Transport in the TTL

Lagrangian trajectory studies are established tools for studying transport processes in the tropical tropopause, and in particular the transport from the troposphere to the stratosphere (*e.g.*, Hatsushika and Yamazaki, 2003, Bonazzola and Haynes, 2004, Fueglistaler *et al.*, 2004). Stratospheric water vapour is strongly correlated with the Lagrangian Cold Point (Fueglistaler *et al.*, 2005). We analyse the minimum temperature ($T_{min}$) and TTL residence time of two CCMVal-2 models, CMAM and E39CA and compare them to ERA-40 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 ERA-40. A 'standard' calculation used 3D winds, and a diabatic calculation used verti-

cal 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 analysed to determine the geographical distribution of points where individual air masses encounter their minimum temperature ($T_{min}$) and thus minimum water vapour 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 analysed, both CCMs have a warm bias in the temperatures of 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 ERA-40 reference calculation. This is not the same as the temperature bias in the models (Figure 7.3). The Eulerian mean tropical *T* is about 3 K low for E39CA and 1 K high for CMAM. Thus, the overall degree of dehydration during transport of air into the stratosphere should 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 7.7** shows that the overall geographical distribution of dehydration points in the simulation based on ERA-40 data are fairly well reproduced by both CCMs in NH winter 1995-1996. 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 vapour flux is shifted eastwards compared to ERA-40 and the model shows excessive water vapour transport through warm regions over Africa. The CMAM model compares very well with the reference calculations, and if anything slightly overestimates the water vapour 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 for 1996, the reference calculations show that the water vapour transport into the stratosphere is clearly dominated by the Indian monsoon and downwind regions (not shown), similar to Fueglistaler *et al*. (2005). This result is largely reproduced by the CMAM model, which also reproduces the location of this feature

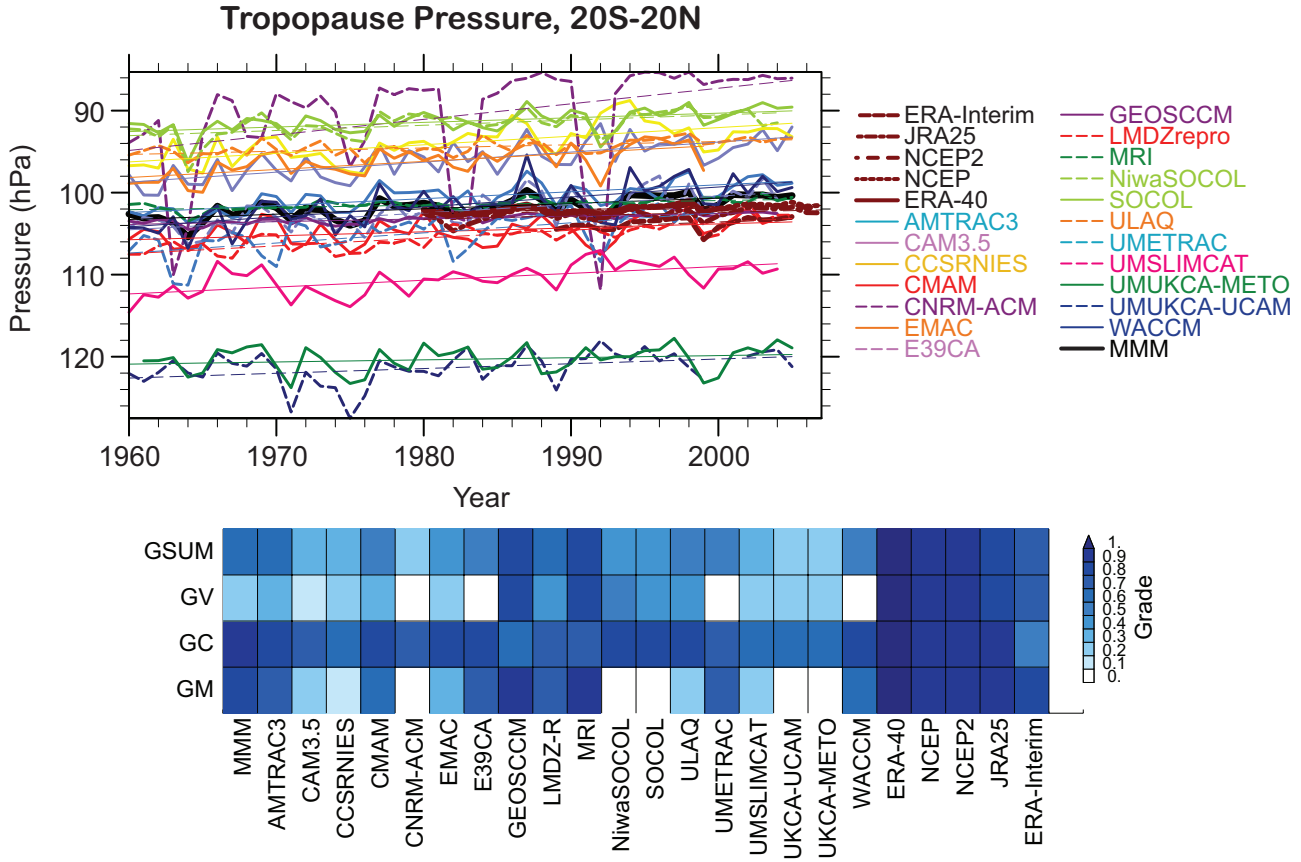


**Figure 7.6:** (Above): Annual mean time series of the Lapse Rate Tropopause Pressure (PTP) for 20°S-20°N from models and analyses for 1960-2007. Thin lines are linear fits. The multi-model mean (MMM) is the thick black line. (Below): Quantitative metrics summary of lapse rate tropopause pressure (PTP) for the mean (GM), the correlation (GC), the variance (GV), and the average of all three grades (GSUM).

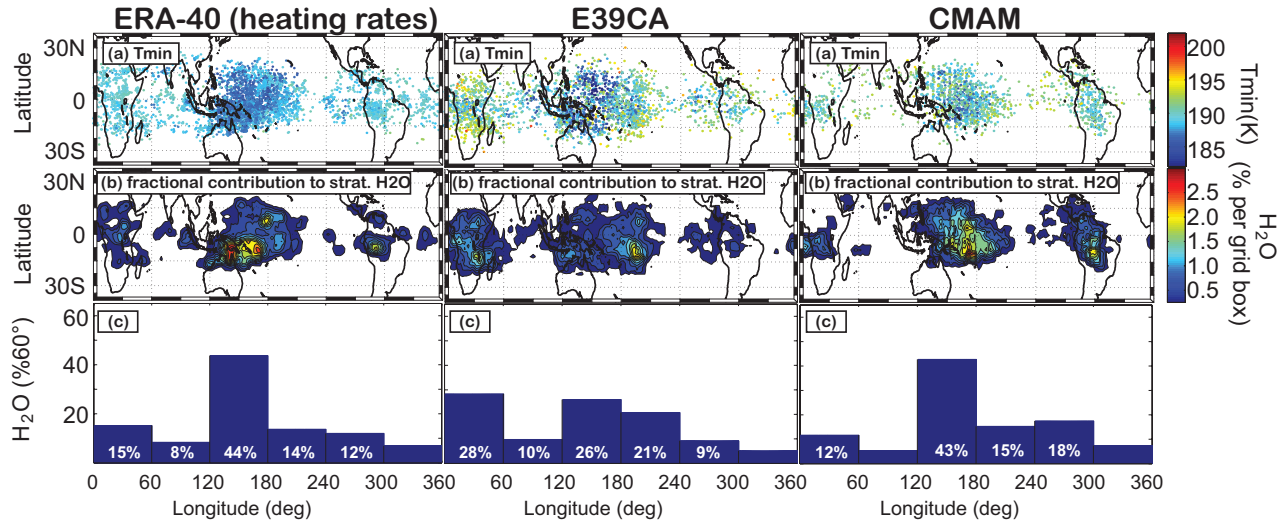


**Figure 7.7:** NH winter 1995-1996. The scatter plots (panel a) show the geographical distribution of the dehydration points for ERA-40 (left), E39CA (middle), and CMAM (right). Colour code in (a) shows the minimum temperatures experienced by the trajectories. Panel (b) illustrates the fractional contribution to stratospheric water vapour from different geographical areas, expressed as percentage contribution per individual 10 x 5 grid boxes (the sum is 100%). Panel (c) shows longitudinal distribution of the water vapor entry value, i.e., the value from (b) integrated over latitude (30ºN-30ºS) per 60º longitude (where the sum is 100%).

nicely. But the water vapour flux through the warm region over Africa is overestimated. In E39CA, the impact of the Indian monsoon is not well reproduced, and dehydration in NH summer for 1996 occurs mostly over the central Pacific rather than over India and the westernmost Pacific.

The residence times in the upper part of the TTL ($\theta$ = 385–395 K) were derived from the trajectory calculations to examine the time scales of transport processes through the TTL. The residence time is a key parameter for chemical transformation of air before it gets into the stratosphere. **Figure 7.8** shows histograms of the probability density function (PDF) of the residence times obtained from the calculations in the upper part of the TTL for NH winter 1995-1996 and NH summer 1996.

Figure 7.8 indicates that on average the reference trajectories stay a few days longer in the upper part of the TTL than the kinematic trajectories, except in NH winter, where the kinematic trajectories calculated with the CMAM data on average stay longer in the TTL than the reference trajectories. In the latter case more kinematic trajectories stay longer than 10 days in the upper part of the TTL than in the reference calculations. The difference between the reference trajectories and the kinematic trajectories is most pronounced based on ERA-40 data, which is consistent with the notion of Schoeberl *et al*. (2003) that assimilation models tends to produce noisier vertical wind fields than free running GCMs. The most important difference between the different panels is the shape of the PDF. The majority of the reference trajectories reside 9-10 days in the upper TTL before they leave this layer. Only very few air masses pass through this layer in less than 5 days. In contrast, for transport that is based on vertical winds (*i.e*., the transport in the CCMs) more trajectories reside 0-5 days in this layer than in the reference calculations, in particular the calculations based on E39CA data result in residence times that are shorter than 5 days for the majority of the trajectories. The percentage of kinematic trajectories based on CMAM data that reside longer than 10 days in the upper part of the TTL is higher than in the reference for both NH winter and NH summer. This is a crucial influence on the transport of short-lived chemical species through the TTL. In E39CA, more short-lived compounds should be able to reach the stratosphere chemically unaltered compared to CMAM and the reference calculation.

The seasonal cycle of the residence time (with slower ascent and longer residence times in NH summer) is represented in the reference calculations (solid bars compared to lines). This is expected from the seasonal variation of the Brewer-Dobson circulation. Both CCMs fail to reproduce the seasonal variation of ascent rates through the TTL.

### 7.4.4 Ozone

The annual cycle of ozone at 100 hPa in the tropics is illustrated in **Figure 7.9**. The annual cycle of $O_3$ near the tropical tropopause is determined by chemical production, vertical transport, and any 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. The

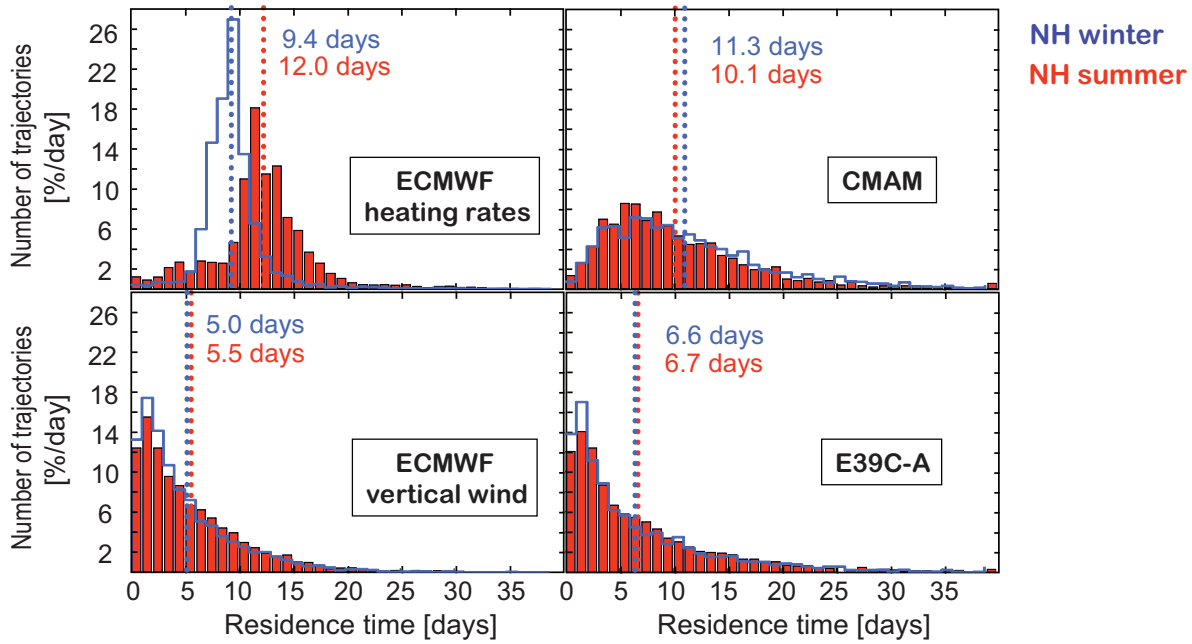


**Figure 7.8:** NH winter 1995-1996 and NH summer 1996. Residence time for the trajectories in the upper part of the TTL (385-395 K) for ERA-40, using heating rates as the vertical velocity (top left), ERA-40 using vertical wind as the vertical velocity (bottom left), CMAM (top right), and E39CA (bottom right). JJA solid red, DJF blue outline. Dotted vertical lines indicate the mean residence time.

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 are based on the annual cycle, variance and anomalies 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. 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.

### 7.4.5 Water Vapour

Water vapour 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 radative impact of water vapour (Kley *et al*., 2000). Thus reproducing the transport of water vapour through the tropical tropopause is a critical requirement of CCMs in the TTL. Representing the appropriate relationships between cold point temperature and water vapour is also critical, as it requires the appropriate representation of processes that regulate water vapour, at least at the large scale.

**Figure 7.10** presents the annual cycle of water vapour from CCMs and HALOE in the lower stratosphere just above the TTL and the cold point (80 hPa). 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 Chapter 5. Here we focus on the entry point. Most models are able to reproduce the annual cycle of water vapour, 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 vapour at this level: from 2-6 ppmv, with observations from HALOE closer to 3-4 ppmv. The uncertainties in HALOE observations are discussed in detail in Kley *et al*. (2000), but are less than ±20% at this level. The shading indicates $3\sigma$ interannual 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 cold point temperature range (Figure 7.3). The multi-model mean indicates that most models shift the water vapour minimum at 80hPa 1-2 months too early. The UMUKCA models fix water vapour in the stratosphere and are not shown. E39CA has too much $H_2O$, consistent with a Lagrangian CPT higher than ERA-40 (Section 7.4.3), but CMAM has too little $H_2O$, despite also having a higher Lagrangian TCPT.

Some models are clearly outside this range, and some have annual cycles that are shifted more than one month, indicating potential problems in transport and dehydration processes. The outliers include MRI (high $H_2O$), CNRM-ACM (high $H_2O$), LMDZrepro (low $H_2O$), and UMETRAC, with virtually no annual cycle (which may be an analysis or data set problem).

Another method of examining the dehydration proc-

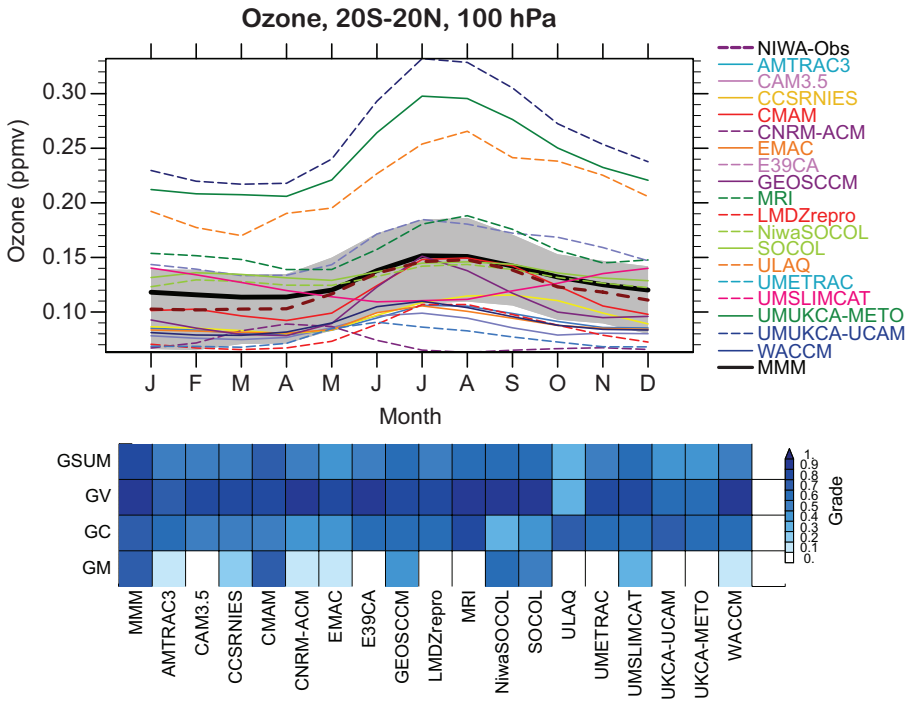


**Figure 7.9:** (Above): Annual cycle of tropical (20°S-20°N) ozone mixing ratio from models and observations. Output and observations are from the period 1980-1999. Gray shaded region is 3σ variability from NIWA observational data set (dashed brown line). (Below): Quantitative metrics summary of 100hPa Ozone for mean (GM), correlation (GC), variance (GV) and the average of all three grades (GSUM).

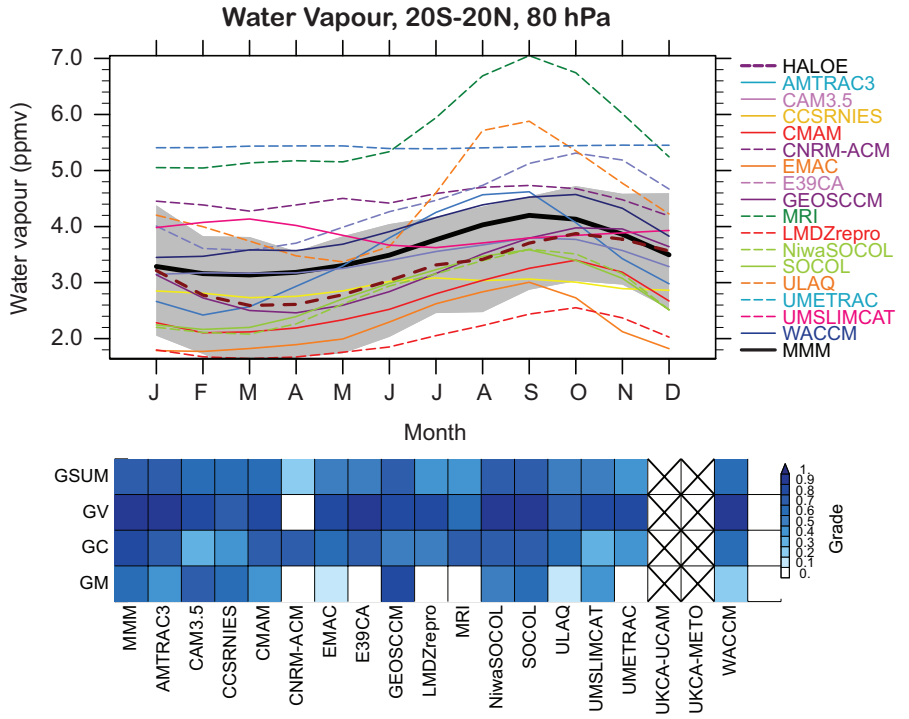


**Figure 7.10:** (Above): Annual cycle of tropical (20°S-20°N) water vapour at 80 hPa from models and observations. Output from the period 1992-2004. Gray shaded region is 3σ variability from HALOE observations over 1992-2004 (brown dashed line). (Below): Quantitative metrics summary of 80hPa $H_2O$ for mean (GM), correlation (GC), variance (GV) and the average of all three grades (GSUM).

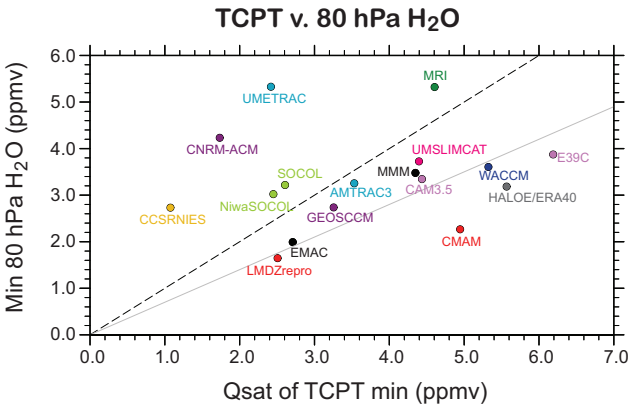


**Figure 7.11:** Correlation of minimum monthly mean water vapour with saturation vapour mixing ratio ($Q_{sat}$) of the minimum monthly mean TCPT from CCMVal-2 models (1980-1999), the multi-model mean (MMM) and HALOE and ERA-40 observations (HALOE over 1992-2005). The black dashed line is the 1:1 line, indicating 100% saturation. The gray line is the 0.7:1 line, indicating 70% saturation.

ess is to look at the relationship between cold point temperature and water vapour just above it. This is a broad way of understanding integrated TTL transport and dehydration in the absence of data for offline Lagrangian cold point calculations as in Section 7.4.3. The TCPT regulates $H_2O$ (Brewer, 1949), so the relationship can be analysed by looking at the ratio of water vapour to the saturation vapour mixing ratio at the cold point ($Q_{SAT}$(CPT)). An update of this relationship, shown in Gettelman *et al*. (2009), is presented in **Figure 7.11**. For example, the minimum ERA-40 TCPT (Figure 7.11) is about 192 K, which corresponds at 80 hPa to a $Q_{SAT}$ of 5.5 ppmv.

Note that the UMUKCA models have very high cold point temperatures (consistent with high ozone at 100hPa and slow transport times), so their water vapour was fixed (and they are not shown). The results indicate that most of the models cluster, similar to the observations ($H_2O$ from HALOE and CPT from ERA-40), near a line that would imply 70% saturation with constant temperatures and transport (which is not the case, hence $H_2O$ is less than implied by TCPT). Three models are near the 1:1 line. MRI is higher than the 1:1 line due to ice-supersaturation permitted in the model. However, three models (CNRM-ACM, CCSRNIES and UMETRAC) have significantly more water vapour than would seem to be justified by their temperatures. This indicates potential problems in fundamental transport, variability and/or condensation processes in the TTL. This is also clear from Figure 7.10.

## 7.4.6 Intra-seasonal Variability/ Waves

There exists significant sub-seasonal variability in temperature and other parameters around the tropical tropopause (*e.g*., Tsuda *et al*., 1994; Fujiwara *et al*., 2009). This is due to equatorial waves, intra-seasonal oscillations (ISOs), and other disturbances that are generated by tropical organised convection (*e.g*., Fujiwara and Takahashi, 2001; Suzuki and Shiotani, 2008). Also, the climatological temperature distribution around the tropical tropopause is in part determined by quasi-stationary disturbances (Highwood and Hoskins, 1998). Therefore, appropriate representation of tropical convection and tropical disturbances is crucial even for stratospheric models, since waves help determine the coldest temperatures, and may affect dehydration in the TTL. For example, with the same minimum $Q_{sat}$ in Figure 7.11, larger wave driven temperature variance would reduce $H_2O$. In this section, the wave activity in temperature at 100 hPa in the tropics is presented for five reanalysis data sets (ERA-40, ERA-Interim, JRA25, NCEP/NCAR (NCEP1), and NCEP-DEO AMIP-II (NCEP2)), and for four CCMs that produced high time frequency winds and temperatures (CCSRNIES, CMAM, MRI, and WACCM), using a zonal-wavenumber-frequency spectral analysis.

All five reanalysis data sets are output four times daily, at a horizontal resolution of 2.5° for ERA-40, NCEP1, and NCEP2, 1.5° for ERA-Interim, and 1.25° for JRA25. CCSRNIES and MRI data are output daily (daily average) at ~2.8° resolution, CMAM data is output four times daily at ~5.6°, and WACCM data is output four times daily at 2.5° by ~1.895°. All data are available for the period between January 1990 and February 2000. All CCM outputs are from the REF-B1 experiment with observed SSTs.

**Figures 7.12** and **7.13** show the zonal-wavenumber-frequency spectrum of temperature at 100 hPa within ~15°N to ~15°S for ERA-40, NCEP1, and four CCMs, for symmetric and antisymmetric components, respectively. Analysis is made for several overlapping 92-day segments for all seasons between January 1990 and February 2000. The spectral calculations include a symmetric-antisymmetric decomposition (Wheeler and Kiladis, 1999). The spectrum shown here is normalised by the variance of the original data, and thus the power spectral density is averaged over the latitude region and the period. The background red-noise spectra, against which the statistical significance is evaluated, are estimated for symmetric and antisymmetric components separately (in frequency only) using the auto-regressive-process method (Gilman *et al*., 1963). Also shown are dispersion curves for theoretical equatorial waves (Matsuno, 1966). Features commonly observed in all data sets are equatorial Kelvin waves (Figure 7.12) and mixed Rossby gravity (MRG) waves (Figure 7.13). ISOs, at frequencies smaller than 0.05 cycle per day, are mostly not statistically significant with respect to the background spectra estimated here; however, the largest power is found in these regions.

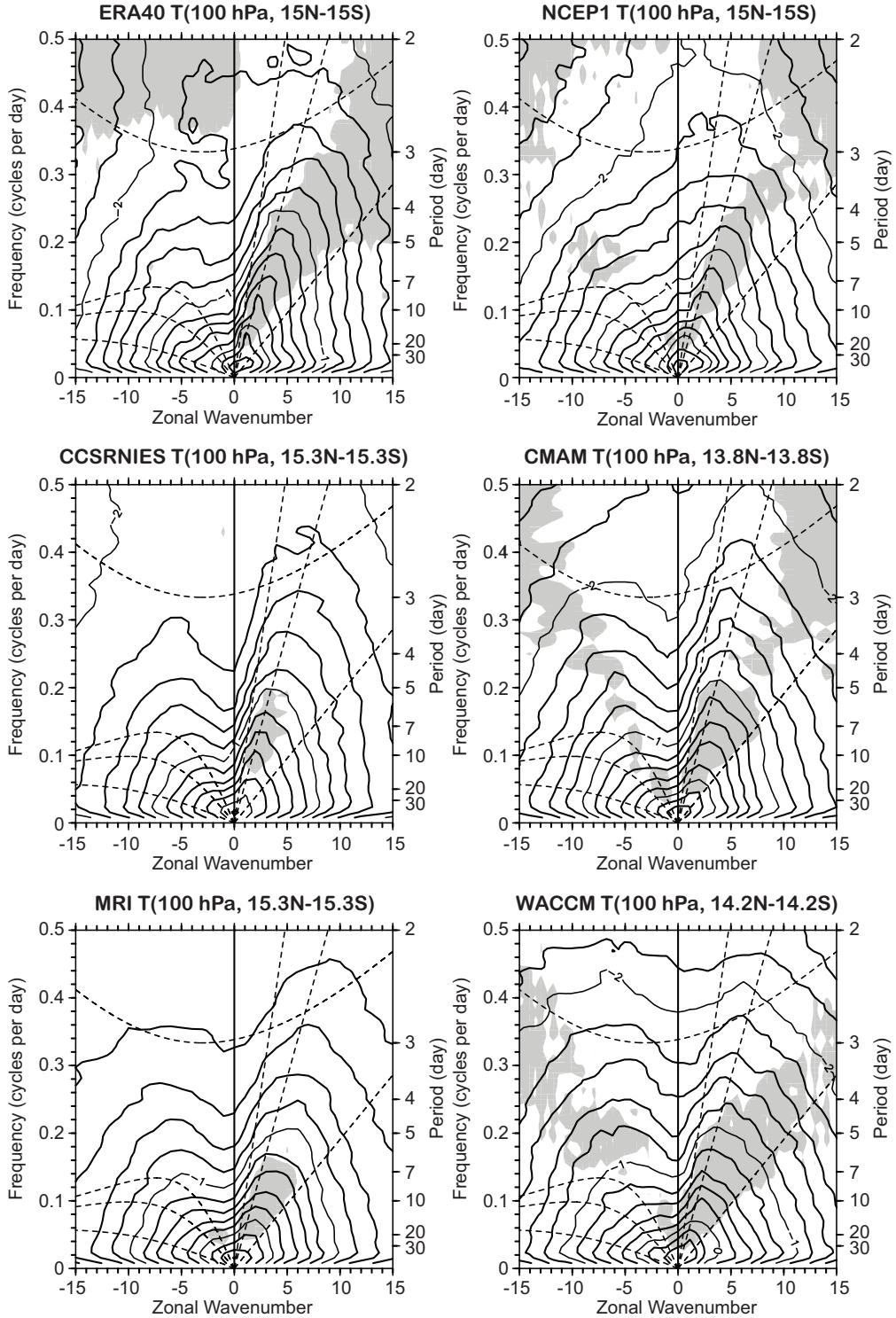


**Figure 7.12:** Zonal-wavenumber-frequency spectrum of temperature at 100 hPa within 15°N-15°S for all seasons between Jan 1990 and Feb 2000 for the symmetric component for ERA-40, NCEP1, CCSRNIES, CMAM, MRI, and WACCM. Contours show the $\log_{10}$ of power spectral density (interval 0.2). Regions where the ratio to the estimated background spectrum is >1.5 are coloured gray. Dotted curves show the wave dispersion relation at equivalent depth, h=8, 70, and 240 m for Kelvin waves (positive wavenumbers) and equatorial Rossby waves (negative wavenumbers). The dispersion relation for meridional-mode-number n=1 inertio-gravity waves with h=8 spans all wavenumbers.

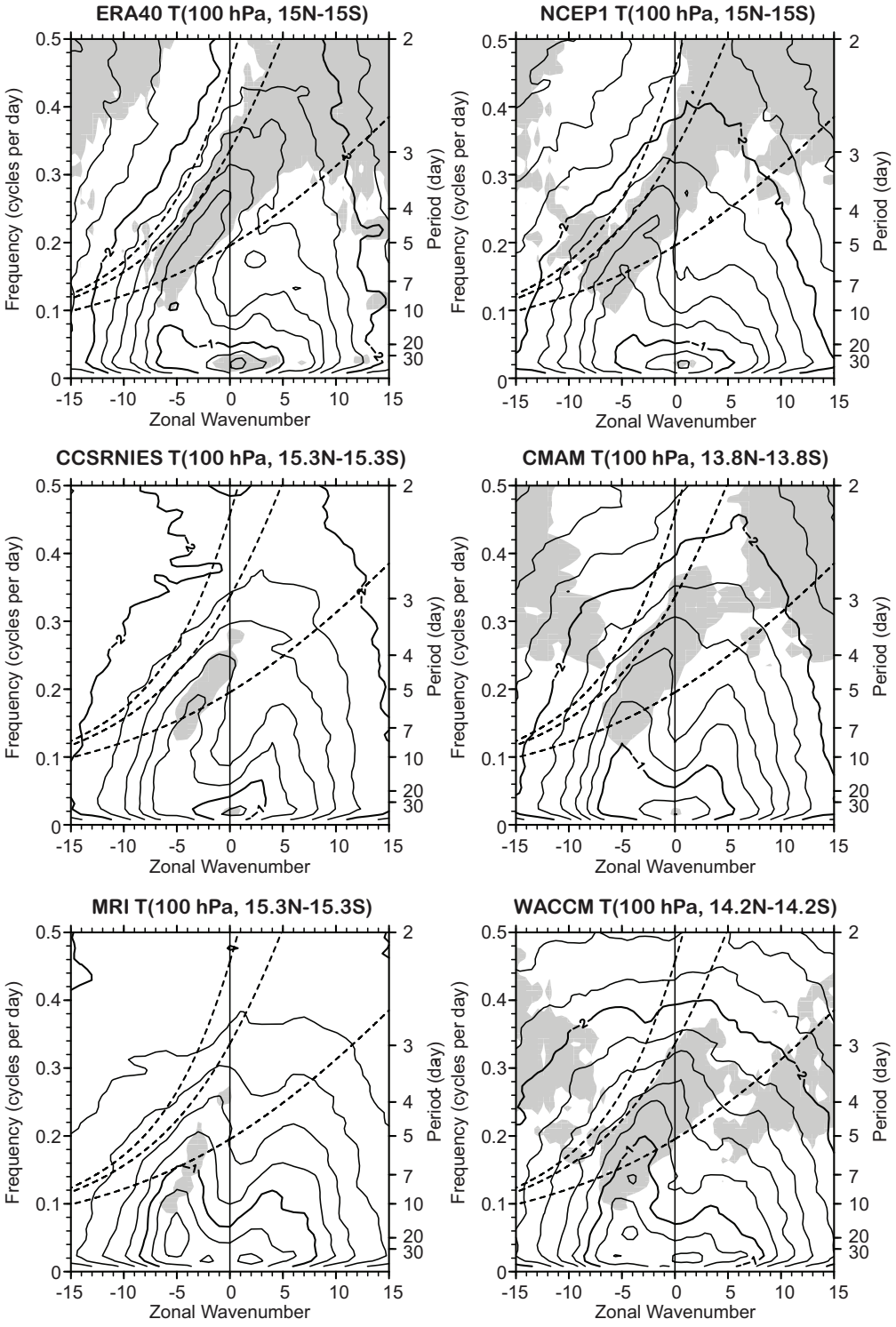


**Figure 7.13:** Same as Figure 7.12 but for the antisymmetric component. Dotted curves show the wave dispersion relation at h = 8, 70, and 240 m for mixed Rossby gravity waves (negative wavenumbers) and n=0 eastward-moving inertio-gravity waves (positive wavenumbers).

**Table 7.2** summarizes the activity for equatorial Kelvin waves, MRG waves, and symmetric eastward-moving ISO. For Kelvin waves, the activity is defined as the integration of power spectral density in the region, zonal wavenumber, $s$ = 1 to 10, frequency, $f$ = 0.05 to 0.5 cycles per day, equivalent depth, $h$ = 8 to 240 m, and the ratio

to the estimated background spectrum ≥ 1.5. For MRG waves, the activity is defined as the integration in the region, s = -10 to 0, $f$ = 0.1 to 0.5 cycles per day, h = 8 to 70 m, and the ratio ≥ 1.5. For the ISO, the activity is defined as the integration in the region, $s$ = 1 to 5, and $f$ = 0 to 0.05; the ratio (the statistical significance) is not considered. Thus, the ISO activity shown here is the upper limit. The activity is then divided by the average activity for five reanalysis data sets for each wave/oscillation.

Results indicate that the five reanalysis data sets have very different wave activities. Fujiwara *et al*. (2009) note that the Kelvin wave amplitudes in the analyses are lower than observed. The Kelvin and MRG wave activities in the ERA reanalyses are 2-3 times greater than those in the NCEP reanalyses; those in JRA25 are close to the average. The ISO activity (the upper limit) is rather similar for the five reanalysis data sets. The wave activities in the four CCMs are generally within the range of those in the reanalysis data. WACCM shows the greatest Kelvin and MRG wave activities, which are comparable to those in the ERA reanalyses. CCSRNIES shows the smallest Kelvin wave activity, and MRI shows the smallest MRG activity. The ISO activity in the CCMs is greater than that in the five reanalysis data sets except for CCSRNIES. Thus the lower range of wave activities is probably too low in these models. The calculation does not explicitly include gravity waves, which may also have significant contributions to temperature variance and dehydration in the TTL.

### 7.4.7 Vertical Thermal Structure

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 so-called "tropopause inversion layer" (TIL) is also characterized by a sharp and strong buoyancy frequency ($N^2=-g/\theta\ d\theta/dz$) maximum. The buoyancy frequency is also called the Brunt-Väisälä frequency. The presence of the TIL has been further confirmed by the Global Positional 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. 7.14a**, d). Although the formation and maintenance mechanisms of the TIL remain to be determined, its presence has potentially important implications for the cross-tropopause exchanges of passive tracers/water vapour and for the dynamical coupling between the stratosphere and troposphere.

**Table 7.2:** The wave activity for equatorial Kelvin waves, mixed Rossby gravity (MRG) waves, and symmetric eastward-moving ISO for five reanalysis and four CCM data sets, with respect to the average value for the five reanalysis data sets (0.184 K² for Kelvin waves, 0.0311 K² for MRG waves, and 0.209 K² for the ISO).

| | Kelvin waves | MRG waves | ISO |
|---|---|---|---|
| ERA-40 | 1.46 | 1.21 | 1.10 |
| ERA-Interim | 1.48 | 1.28 | 1.03 |
| JRA25 | 1.12 | 0.997 | 1.22 |
| NCEP1 | 0.524 | 0.826 | 0.859 |
| NCEP2 | 0.421 | 0.684 | 0.786 |
| CCSRNIES | 0.344 | 0.416 | 0.716 |
| CMAM | 1.22 | 0.799 | 1.34 |
| MRI | 1.04 | 0.234 | 1.50 |
| WACCM | 1.44 | 1.43 | 1.31 |

The zonal-mean structure of the TIL, simulated by REF-B1 integrations for 9 models with available instantaneous data, is examined and compared with observations. The observed TIL is derived from the COSMIC GPS RO data set (Anthes *et al*., 2008).

All analyses are performed on the log-pressure coordinate with tropopause pressure ($p_{TRP}$) as a reference level: *i.e*., $z = H \ln (p/p_{TRP})$ where $H$ is a scale height of 8 km. Note that the conventional log-pressure coordinate uses surface pressure for a reference level. At each model grid point (or GPS RO profile) tropopause pressure is first computed 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-pressure linear interpolation, and are averaged over longitudes for December-January-February (DJF) and June-July-August (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 (Figure 7.14a, d) and data only at CCMVal-2 standard levels (Figure 7.14b, e). CCMVal-2 UTLS standard levels are shown in Figure 7.5. Degraded observations reduce uncertainties associated with model vertical resolution, and allow a more direct comparison of the simulated TIL with observations.

The analysis results for the average of 9 models are summarized in Figure 7.14 in terms of $N^2$. We first describe the TIL in the observations. As shown in Figures 7.14a, d, sharp maxima of $N^2$, located just above the tropopause ($z$ = 0), are distinct in the extra-tropics. They are generally stronger in the summer hemisphere than in the winter hemisphere, but have no hemispheric difference: *i.e*., the $N^2$ distribution in the NH summer is quantitatively similar

to the one in the SH summer. In contrast, the tropical $N^2$ profile is only weakly sensitive to season. This is consistent with previous findings (Randel *et al*., 2007; Grise *et al*., 2009).

Figures 7.14b, e show the $N^2$ distribution for degraded GPS RO data. Maximum values of $N^2$ are substantially weakened. In addition, their locations are somewhat higher than those in the raw data. 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 under-estimated in coarse resolution GPS RO 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. 7.14c, f) is gener-

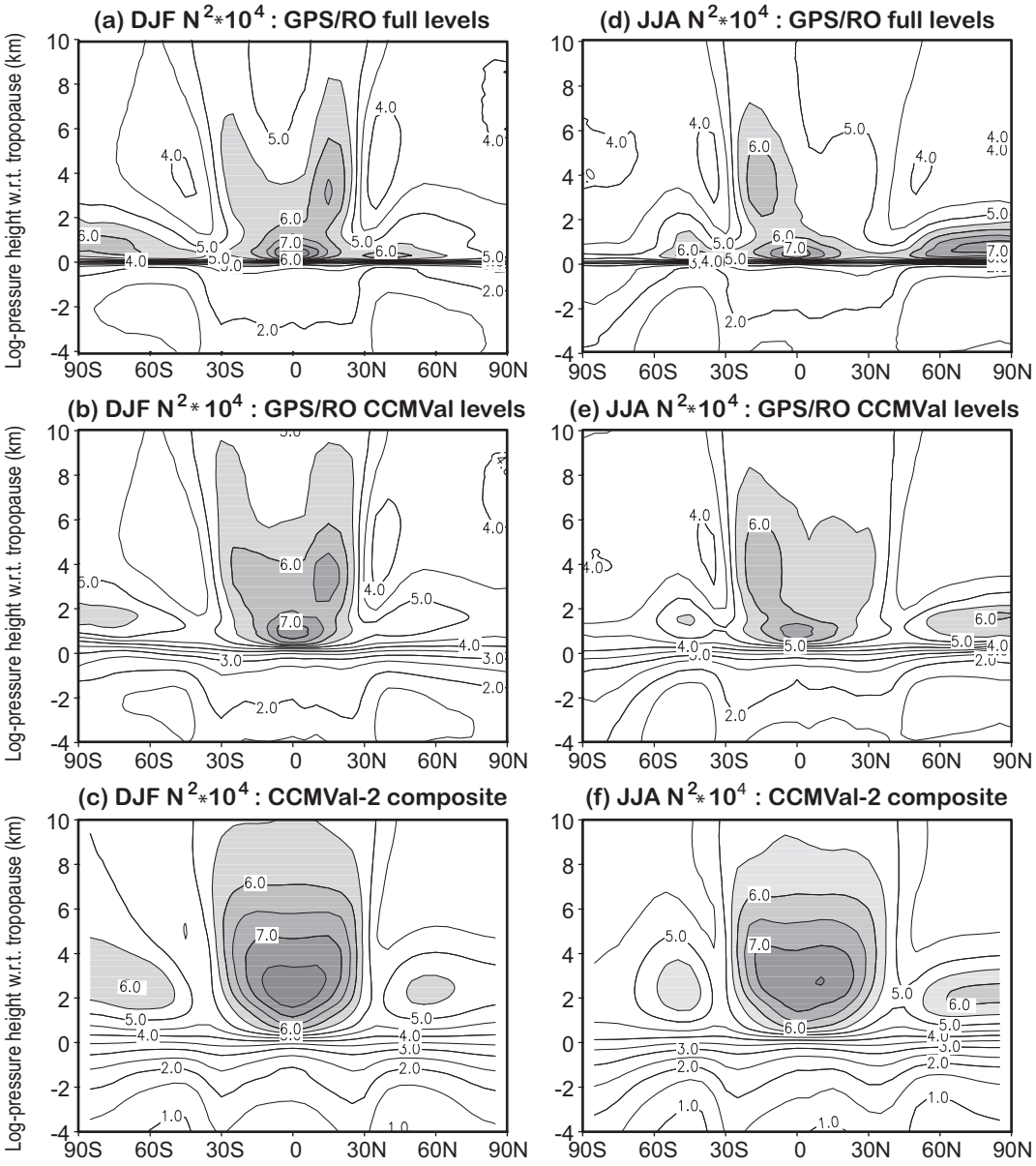


**Figure 7.14:** Zonally-averaged $N^2$ ($\times 10^4$) as a function of latitudes and log-pressure height on the tropopause based coordinate: (top) COSMIC/FORMOSAT-3 GPS RO data, (middle) COSMIC GPS RO data using only CCMVal-2 standard pressure levels, and (bottom) composite of 9 REF-B1 model integrations. Two seasons are shown separately: (left) December-January-February and (right) June-July-August. Contour intervals are 0.5 $s^{-2}$, and values greater than or equal to 5.5 $s^{-2}$ are shaded. Note that zero in y-axis denotes the location of the tropopause.

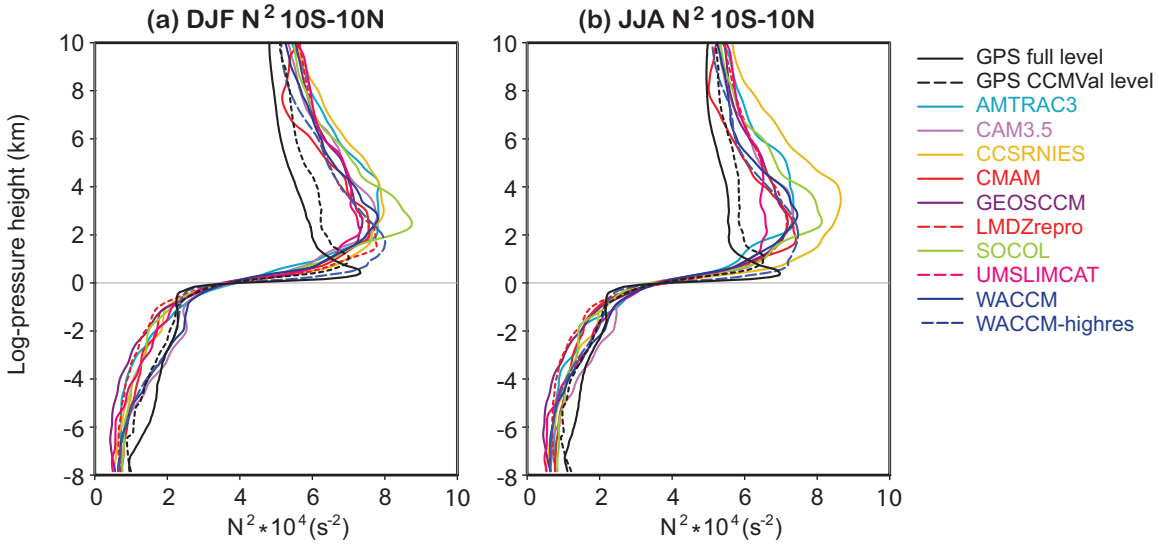


**Figure 7.15:** Vertical profiles of $N^2$ ($\times 10^4$) in each model and observation in the tropics. (a) DJF, (b) JJA.

ally weaker and broader than observed using full resolution data (Figures. 7.14a, d). The simulations do look more like estimates from the observations using CCMVal vertical resolution. (Figures 7.14b, e). Analysis of higher vertical resolution runs from WACCM with 300 m vertical resolution in the UTLS (WACCM-highres) does indicate that at a higher vertical resolution this model has an increased peak $N^2$ near the tropopause in better agreement with GPS RO observations (Figure 7.15, also Figure 7.19).

**Figure 7.15** illustrates profiles of $N^2$ from GPS RO observations and simulations in the tropics for 2 seasons from 9 models and WACCM-highres. The CCMVal-2 models under-estimate $N^2$ in the troposphere and misplace the tropical temperature inversion layer. 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 under-estimated adiabatic cooling from tropical upwelling and/or radiative cooling associated with lower stratospheric $H_2O$. Note that WACCM-highres has a larger and sharper peak in $N^2$, and the peak is closer to the tropopause than the standard resolution WACCM. In addition, two of the lower vertical resolution models analysed (CCSRNIES and SOCOL) have very broad TIL structures. The discussion on the extra-tropical TIL is continued in Section 7.5.1.4.

## 7.5 The Extra-tropical UTLS

The extra-tropical UTLS is here defined as the region between the free troposphere (6-8 km) and the upper boundary of the tropically controlled transition region (around 22 km, Rosenlof *et al*., 1997) as illustrated in Figure 7.1. It includes the Lowermost Stratosphere (LMS), the region between the extra-tropical tropopause and the 380 K potential temperature surface (Holton *et al*., 1995). One main characteristic of the LMS is that isentropes intersect the tropopause, thereby potentially connecting the troposphere and the stratosphere *via* rapid adiabatic motion. The slower diabatic circulation is predominantly downward in the LMS, which on its own would transport aged stratospheric air into this region. However, meridional mixing from the tropical UTLS transports younger air masses to mid- and high latitudes and 'rejuvenates' air as it slowly descends into the LMS (Rosenlof *et al*., 1997; Bregman *et al*., 2000; Hegglin and Shepherd, 2007), an effect quantified by Hoor *et al*. (2005) and Bönisch *et al*. (2009) based on SPURT aircraft data, and Levine *et al*. (2007, 2008) using operational analyses of the European Centre for Medium-range Weather Forecasts (ECMWF). The lower boundary of the LMS is defined by the tropopause. Distributions of chemical tracers that are affected by transport exhibit strong spatial gradients across the tropopause in a layer of finite depth referred to as the Extra-tropical Tropopause Transition Layer (ExTL) (Fischer *et al*., 2000; Zahn *et al*., 2000; Hoor *et al*., 2002, 2004; Pan *et al*., 2004). The ExTL is a global feature with increasing depth towards high latitudes, and has been found to be different for different tracers (Hegglin *et al*., 2009). The ExTL chemical transition has been interpreted as the result of recurrent wave-breaking events, forced by synoptic-scale baroclinic disturbances, which bring tropospheric and stratospheric air masses with very different chemical and radiative characteristics into close proximity (Shepherd, 2007). Indeed, Berthet *et al*. (2007) found an analogue of the ExTL using large-scale trajectories driven by ECMWF wind fields. Small-scale processes such as three-dimensional turbulence and ultimately molecular diffusion then act to reduce the gradients produced in the tracer fields (Hegglin *et al*., 2005).

The extra-tropical UTLS is very sensitive to climate

change and will cause chemical, radiative and dynamical feedbacks, due to high sensitivity to changes in the UTLS. Changes in the extra-tropical UTLS help determine the stratospheric impact on the troposphere through *e.g.*, the transport of stratospheric ozone into the troposphere or surface UV fluxes. Thus it is important that CCMs are capable of resolving chemical and dynamical structures in the extra-tropical UTLS accurately. Here we will investigate the CCMs' capability to reproduce the complex dynamical and chemical structures of the extra-tropical UTLS. Potential long-term changes in these structures will be investigated in Section 7.6.2.

*Detailed description of extra-tropical diagnostics*

The following diagnostics are used to obtain performance metrics for the CCMs:

- ***Diagnostic 1:*** The seasonal zonal-mean zonal wind is used to test the models' realism in representing the latitudinal gradients of the thermal structure.
- ***Diagnostic 2:*** The seasonal cycle in the LMS mass is a test of the combined radiative-dynamical response to radiative forcing. It can be seen as an integrated measure for the extra-tropical tropopause behaviour, which is a basic measure of the UTLS thermal structure in a model.
- ***Diagnostic 3:*** The seasonal cycles in $O_3$, $HNO_3$, and $H_2O$ at 100 and 200 hPa are used to test the models' representation of the large-scale transport and mixing properties. This includes the evaluation of the representation of the seasonal relative strength in quasi-horizontal mixing between the tropical latitudes and the extra-tropics within the tropically controlled transition region (380-420 K, or ~100 hPa) and across the subtropical jet (340-380 K, or ~200 hPa). $HNO_3$ in addition is a tracer not only influenced by transport, but also by more complex microphysical and chemical processes (a topic that clearly needs to be addressed more thoroughly in the future).
- ***Diagnostic 4:*** The sharpness of the meridional gradients of long-lived species (here for $O_3$), where long-lived has to be seen in relation to the transport time scales, is a measure of the chemical distinctiveness of the UTLS in latitude, and therefore for the degree of isolation of different regions such as the tropics and the extra-tropics.
- ***Diagnostic 5:*** Vertical profiles of normalised CO in potential temperature units relative to the tropopause height allows us to separate between transport across the extra-tropical tropopause on short time scales and transport from the tropics and subtropics on longer time scales. It thereby helps to determine the tropospheric influence on the lowermost stratospheric background. The normalisation ensures that the results are dependent purely on transport and mixing processes, and not on the boundary conditions of CO in the troposphere.
- ***Diagnostic 6:*** A basic test of the models' performance in the UTLS region uses annual and seasonal profiles of $H_2O$, CO, and $O_3$ in tropopause coordinates at mid-latitudes and northern hemisphere polar regions. This diagnostic is critical for understanding the chemical structure (including sources and sinks) of, and the separation between the UT and LS.

Diagnostics not used in a quantitative way are:

- ***Diagnostic 7:*** Interannual anomalies in extra-tropical tropopause pressure are a measure of the response of the models to different forcings such as volcanoes, ENSO, *etc*. The anomalies are related to LMS mass.
- ***Diagnostic 8:*** The tropopause inversion layer (TIL) is a distinctive feature of the thermal structure of the tropopause, which reflects the balance between radiative and dynamical processes.
- ***Diagnostic 9:*** The depth of the extra-tropical tropopause transition layer (ExTL) and its location relative to the thermal tropopause are used to diagnose the mixing and transport characteristics of the models in the tropopause region.
- ***Diagnostic 10:*** Ozone probability density functions are used to test the variability of ozone with respect to the tropopause.

## 7.5.1 Dynamical Structure of the Extra-tropical UTLS

### 7.5.1.1 Zonal mean wind

The zonal mean zonal wind field is a very common diagnostic and used to validate the representation of the latitudinal thermal structure of the models, and therefore the basic dynamical state of the models' atmospheres. For this diagnostic, monthly zonal mean wind fields averaged over the period 1979-1999 are compared between the REF-B1 simulations and ERA-40. For further comparison, also NCEP data are included.

**Figures 7.16** and **7.17** (which lists the grade for the mean $g_m$ and the skill $S$ calculated using Equations (7.1) and (7.7), respectively, as well as the total grade $G_{tot}$ calculated using Equation (7.8)) illustrate that the models represent the strength and latitudinal behaviour of the zonal-mean zonal wind in a realistic way. This is to be expected since the models usually tune their gravity wave parameterizations towards getting the observed zonal-mean wind fields correct. ULAQ is the only model that shows clear deficiencies in resolving the latitudinal structure, especially during JJA. This lack of realism is also expressed in the Taylor

diagrams by very low (latitude-by-latitude) correlation and skill values, and might be attributable to the very low resolution of the model and its quasi-geostrophic dynamical core. The grades of the mean values of the zonal-mean zonal wind, $g_m$, also reveal that SOCOL-models score slightly lower than the multi-model mean during both DJF and JJA.

The tight correspondence between NCEP and ERA-40 (the skill of NCEP is 0.98), which is tighter than the model spread, indicates good agreement between the two reanalyses, and that the models may still have room for improvement. For example, several models displace the tropospheric 'eddy-driven' jet in the SH summer (DJF) when compared to the observations.

The total grading values ($G_{tot}$) in Figure 7.17 are averaged and listed in the final grading Figure 7.39.

### 7.5.1.2 Mass of the Lowermost Stratosphere

The seasonal cycle in the LMS mass is a basic test of the combined radiative-dynamical response to radiative forcing and represents an integrated measure for the extratropical tropopause behaviour. Stratospheric mass variations due to seasonal tropopause height variations can contribute to stratosphere-troposphere exchange (Appenzeller *et al*., 1996). This exchange transports ozone (*c.f.* Chapter 10) and reactive nitrogen (besides other species) into the troposphere, where it helps determine the tropospheric ozone budget and hence air quality. Here, we test the realism of the seasonal fluctuations in the total LMS mass by comparing them to the NCEP reanalyses using a method similar to the one of Appenzeller *et al*. (1996). The LMS mass is determined as the fraction of the stratosphere that lies between the thermal tropopause, calculated using the WMO definition, and the 100 hPa pressure surface. The thermal tropopause is derived from monthly zonal mean temperature fields averaged over a time period between 1990 and 1999 using the REF-B1 simulations. The results are shown in **Figure 7.18**.

In the NH, most models show a very high skill (with values larger than 0.9) in reproducing the amplitude and phase of the seasonal cycle in the LMS mass from the NCEP reanalyses. One exception is LMDZrepro which scores lower with a value of 0.62. LMDZrepro captures the structure of the seasonal evolution (expressed in a seasonal correlation of 0.95), but under-estimates its amplitude (expressed in a normalised standard deviation of 0.5). There are also quite a few models that have difficulty in simulating accurate mean values of the LMS mass as shown in **Figure 7.19**. UMUKCA-METO and UMUKCA-UCAM show larger LMS mass values, indicating an average tropopause pressure that is too high. CCSRNIES, CNRM-ACM, EMAC, NiwaSOCOL, SOCOL, and ULAQ have smaller LMS mass values than expected, indicating generally too low tropopause pressures. The multi-model mean shows both a good mean value and a high skill comparable to those values obtained by the best performing models AMTRAC3, CMAM, GEOSCCM, and E39CA.

In the SH, the models' overall performance relative to NCEP is worse than in the NH. The skill based on the correlative metrics lies around 20-40% lower than in the NH for all models, with particular deficiencies for CAM3.5, CCSRNIES, EMAC, GEOSCCM, LMDZrepro, ULAQ, UMSLIMCAT, and WACCM. The Taylor diagram reveals

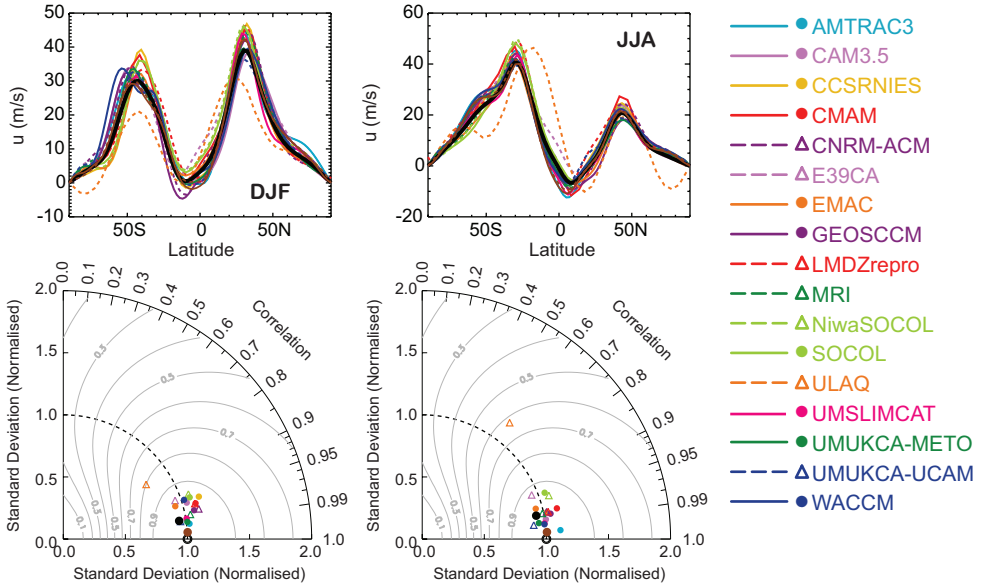


**Figure 7.16:** Zonal-mean zonal wind (upper panels) and corresponding Taylor diagrams (lower panels) at 200 hPa for DJF (left panels) and JJA (right panels). The brown solid line represents ERA-40 data, the brown dashed line and brown dot diagram NCEP data, and the black solid line and dot the multi-model mean.

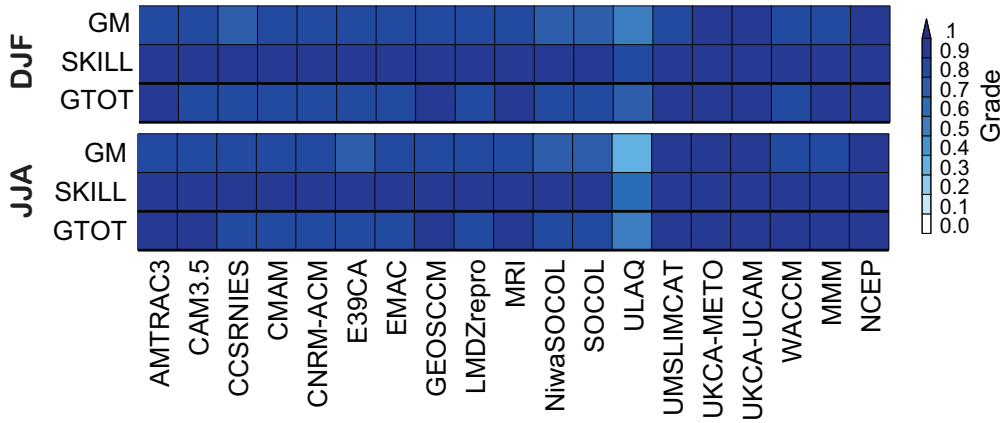


**Figure 7.17:** Grades for the zonal-mean zonal wind at 200 hPa for DJF (upper) and JJA (lower). MMM indicates the multi-model mean. GM is calculated using Equation (7.1), SKILL (which is equivalent to S) using Equation (7.7), and GTOT using Equation (7.8).

that almost all models exhibit standard deviations that are too large, which shifts them away from the reference point (note the different radial axis scale in the Taylor diagrams in Figure 7.18). The following models have major deficiencies in representing the mean values (see Figure 7.19): CNRM-ACM and the UMUKCA models.

An acceptable total score which is equal to that of the multi-model mean (≥ 0.8) is only reached by E39CA. The difference between the SH and NH can be explained by smaller seasonal variations in the LMS mass in the SH, which is more difficult for the models to capture.

The total grading values ($G_{tot}$) obtained in Figure 7.19 for the NH and SH are averaged atnd listed in the final grading Figure 7.39.

### 7.5.1.3 Extra-tropical Tropopause pressure

The extra-tropical tropopause pressure is a basic measure of the thermal structure in a model. We here focus on interannual anomalies in tropopause pressure that yield insight into the models' abilities to respond to forcing of the climate system. The tropopause is calculated using the WMO-definition and averaged over a year and 40°N-60°N and 40°S-60°S, respectively. The analysis is based on

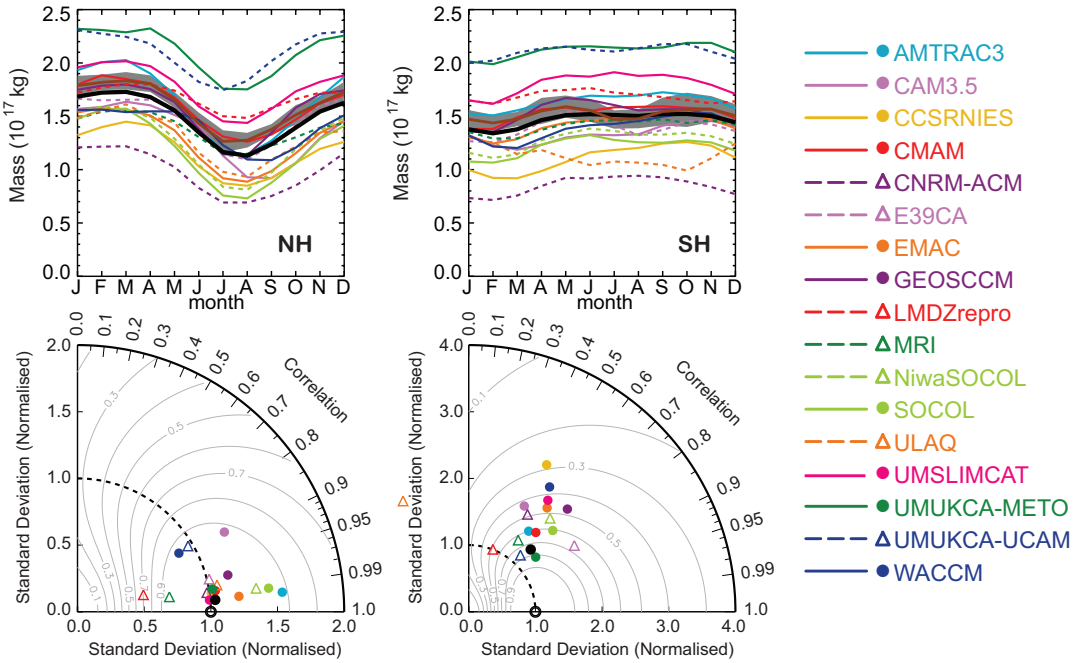


**Figure 7.18:** Seasonal cycle in LMS mass following Appenzeller et al. (1996) (upper panels), and corresponding Taylor diagrams of model performance (lower panels) for NH (left panels) and SH (right panels). Coloured lines, dots and triangles denote models, black solid line and dot the multi-model mean, and brown line and gray shading the NCEP reanalyses ±1σ standard deviation.

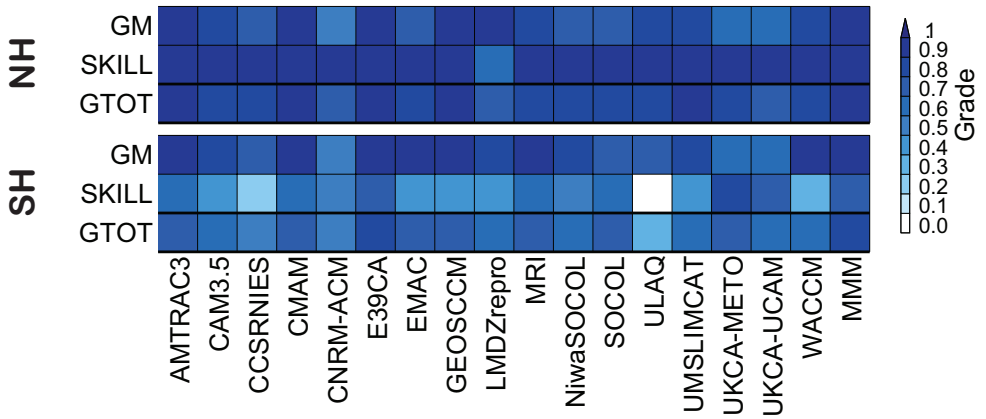


**Table 7.19:** Same as Figure 7.17, but for LMS mass in the NH (upper) and SH (lower).

monthly mean temperature fields (T3M) from the REF-B1 runs. The models are compared to 5 different analyses, ERA-Interim, ERA-40, NCEP, NCEP2, and JRA25.

Although the models seem to reproduce the seasonal cycle in tropopause pressure well in the NH (which can be argued based on the diagnostic for the LMS mass), they show more problems in representing interannual variability. This can be seen from **Figure 7.20**. Similar to the evaluation in the tropics, CNRM-ACM has unrealistic interannual variability and low tropopause pressure. CCSRNIES, EMAC, ULAQ and WACCM achieve lowest total scores (not shown) due to both too high/low mean values and smaller correlation with the observed variability structure.

In the SH, the models simulate the interannual variability somewhat better, except CNRM-ACM which has large interannual variability and as in the NH a too low tropopause pressure. CCSRNIES, MRI, ULAQ, and WACCM have a negative bias in the mean tropopause pressure, and ULAQ shows the worst correlative score.

### 7.5.1.4 Extra-tropical Tropopause Inversion Layer

**Figure 7.21** shows the $N^2$ profiles at two latitude bands, representing the NH TIL in winter and summer (for discussion of full cross-section see Section 7.4.7). It can be seen that maximum values of simulated $N^2$ are comparable to or larger than those derived from degraded GPS RO data. However, they are always weaker than those computed from full-level GPS RO data unless vertical resolution is sufficiently high (*e.g.*,WACCM-highres). It is also evident that the location of maximum $N^2$ in the CCMVal-2 models is always higher above the tropopause than in ob-

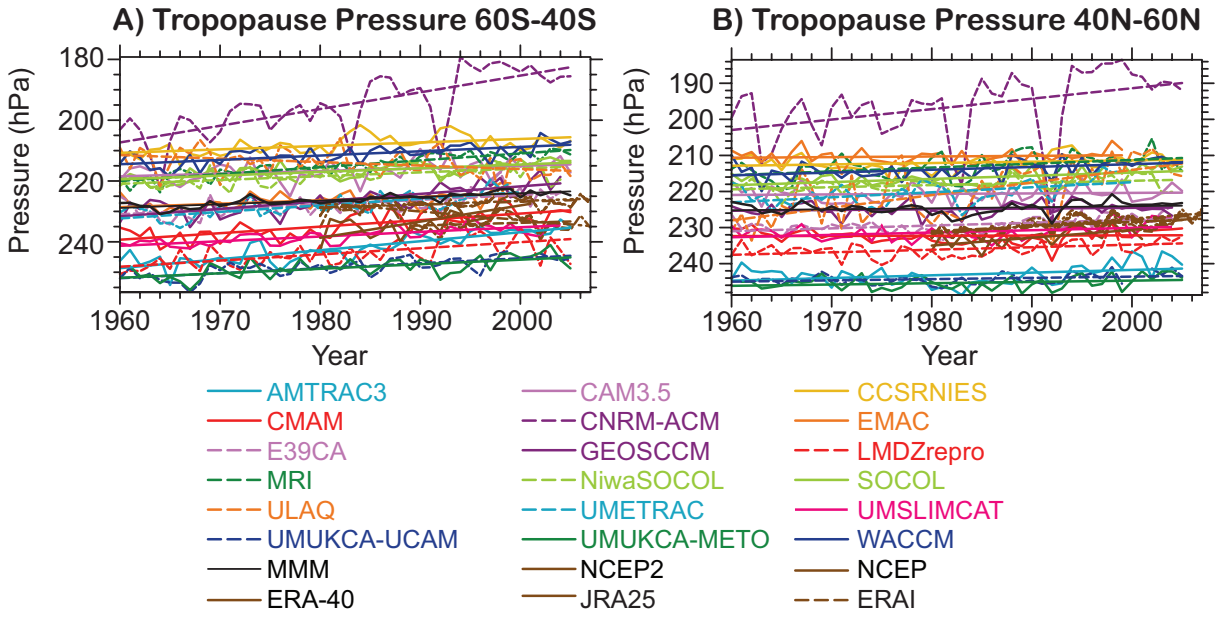


**Figure 7.20:** Extra-tropical tropopause pressure variability for SH (left panel) and NH (right panel).

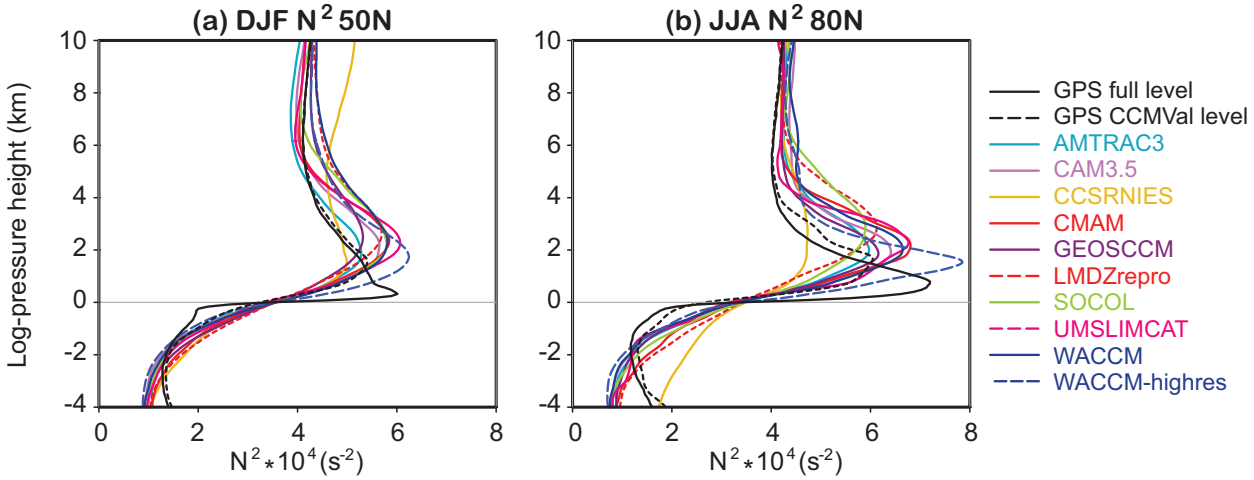


**Figure 7.21:** Vertical profiles of $N^2$ ($\times 10^4$) in models and observation at (a) 50°N during DJF and (b) 80°N during JJA.

servations. These results are consistent with the findings in the tropics (Figure 7.15) and those by Bell and Geller (2008) as discussed in Section 7.4.7.

It should be emphasised that, although the maximum $N^2$ values of the TIL are somewhat under-estimated, most CCMVal-2 models qualitatively reproduce the seasonal and latitudinal changes observed in the TIL. In fact, the models' simulated TIL is more realistic than that derived from reanalysis data (Birner *et al.*, 2006). This may be because the reanalysis systems are ingesting data that may cause degradation to the structure, either through errors or through coarse vertical resolution (*e.g.*, satellite temperatures).

## 7.5.2 Transport and mixing

The chemical structure of the extra-tropical UTLS and its seasonal evolution is determined by the source/sink characteristics of the various species, together with the relative strength of large-scale and small-scale transport and mixing processes. The tracers we focus on in this report ($O_3$, $H_2O$, $HNO_3$, and CO) are relatively long-lived compared to the transport time scales determining their distributions across the extra-tropical UTLS, therefore, the chemical structure of these tracers can be used to validate the underlying transport processes. While large-scale and small-scale mixing processes are hard to disentangle completely, their relative importance is strongly dependent on the sub-region one is considering.

### 7.5.2.1 Tracer seasonal cycles in the 'background' Lowermost Stratosphere

The large-scale Brewer-Dobson Circulation (BDC), driven by (planetary, gravity, and synoptic-scale) wave-drag in the stratosphere, transports aged stratospheric air into the LMS (Logan, 1999). The breaking of synoptic scale waves above the subtropical jet mixes younger tropical air masses with older higher-latitude air masses. The BDC and synoptic-wave transport exhibit a seasonally varying strength, and determines the chemical background composition of the LMS (Hoor *et al.*, 2005; Hegglin and Shepherd, 2007). It is crucial for CCMs to capture the relative strength and seasonality of these processes. This is because they determine the distribution of the radiatively active species $O_3$ and $H_2O$, which through radiative heating can alter temperature distributions and thereby winds in the UTLS, and also determine the monthly input of stratospheric ozone into the troposphere. Lowermost stratospheric background $O_3$ furthermore determines the impact of aircraft emissions on ozone at these altitudes. $H_2O$ plays an important role as precursor of $HO_x$ ($OH + HO_2$) which are the dominant radicals for ozone destruction in the LMS.

The models' representation of these large-scale transport and mixing processes, with typical time scales of weeks to a couple of months, is evaluated here using the seasonal cycles in $O_3$, $HNO_3$, and $H_2O$ at 100 and 200 hPa for latitude bands between 40° and 60°N/S, respectively. While $O_3$ and $HNO_3$ are expected to yield about the same seasonal cycles since their sources are mostly stratospheric at these levels, $H_2O$ is a tropospheric tracer (since the contribution of $CH_4$ oxidation to total water is small) and gives insight into a possible tropospheric influence as well as the lowest saturation vapour pressure an air parcel has experienced. $HNO_3$ is further affected by chemistry and microphysics, which may cause some differences in its seasonal cycle when comparing it to that of $O_3$. The monthly mean zonal-mean tracer fields of the REF-B1 simulations from 2000-2006 are compared to observations obtained by the MIPAS instrument between 2004 and 2008.

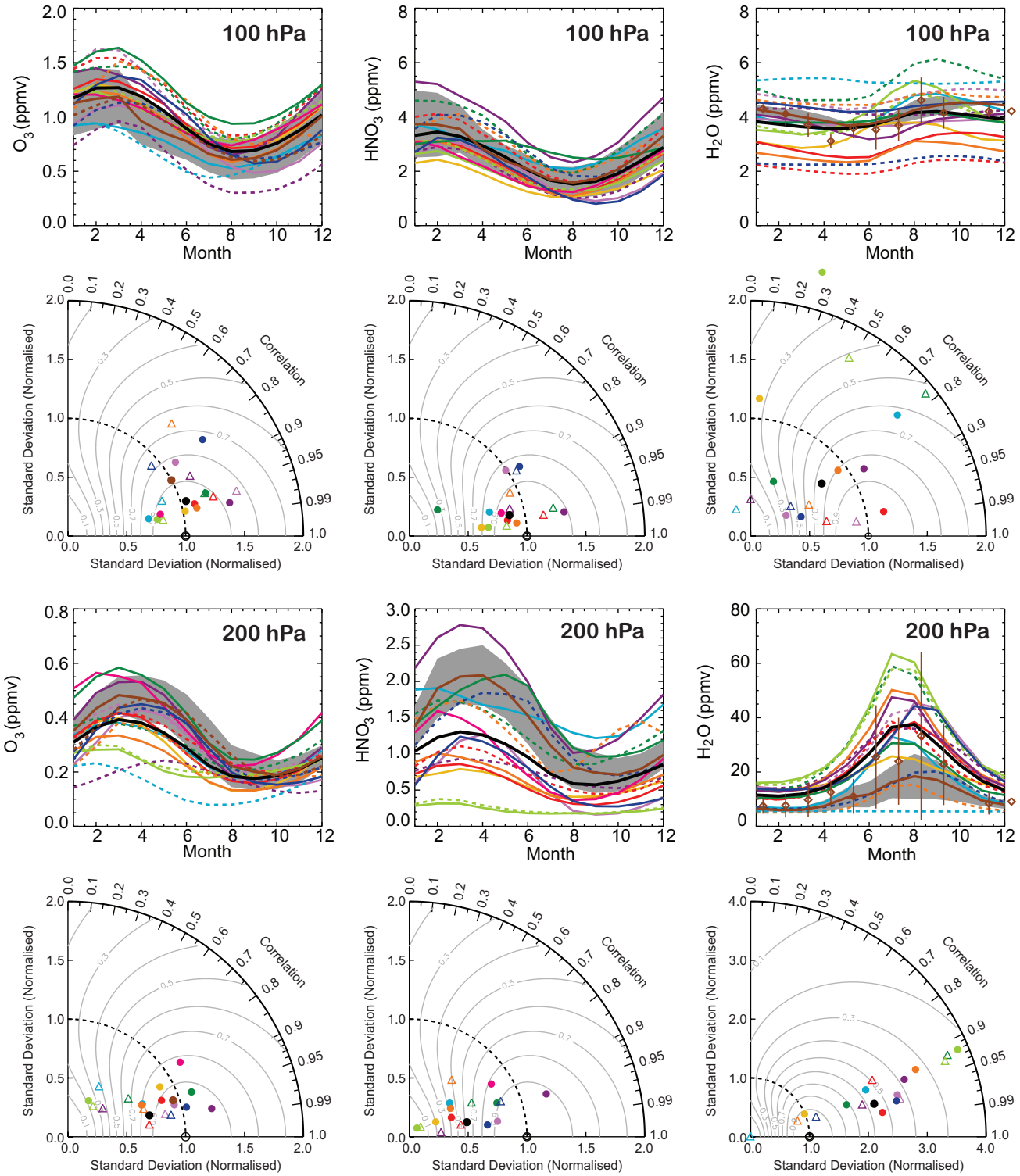


**Figure 7.22:** Seasonal cycles in monthly mean $O_3$, $HNO_3$, and $H_2O$ between 40°N and 60°N and corresponding Taylor diagrams at 100 hPa (upper two rows) and 200 hPa (lower two rows) for different models (colour-coded) compared to MIPAS satellite data and their 1σ uncertainty (brown solid lines and gray shading) over the years 2004-2008. For $O_3$, MLS data over the years 2004-2008 (brown dashed lines and dots), and for $H_2O$, ACE-FTS data (brown diamonds) are shown in addition to the MIPAS data. Black lines and dots indicate the multi-model mean.

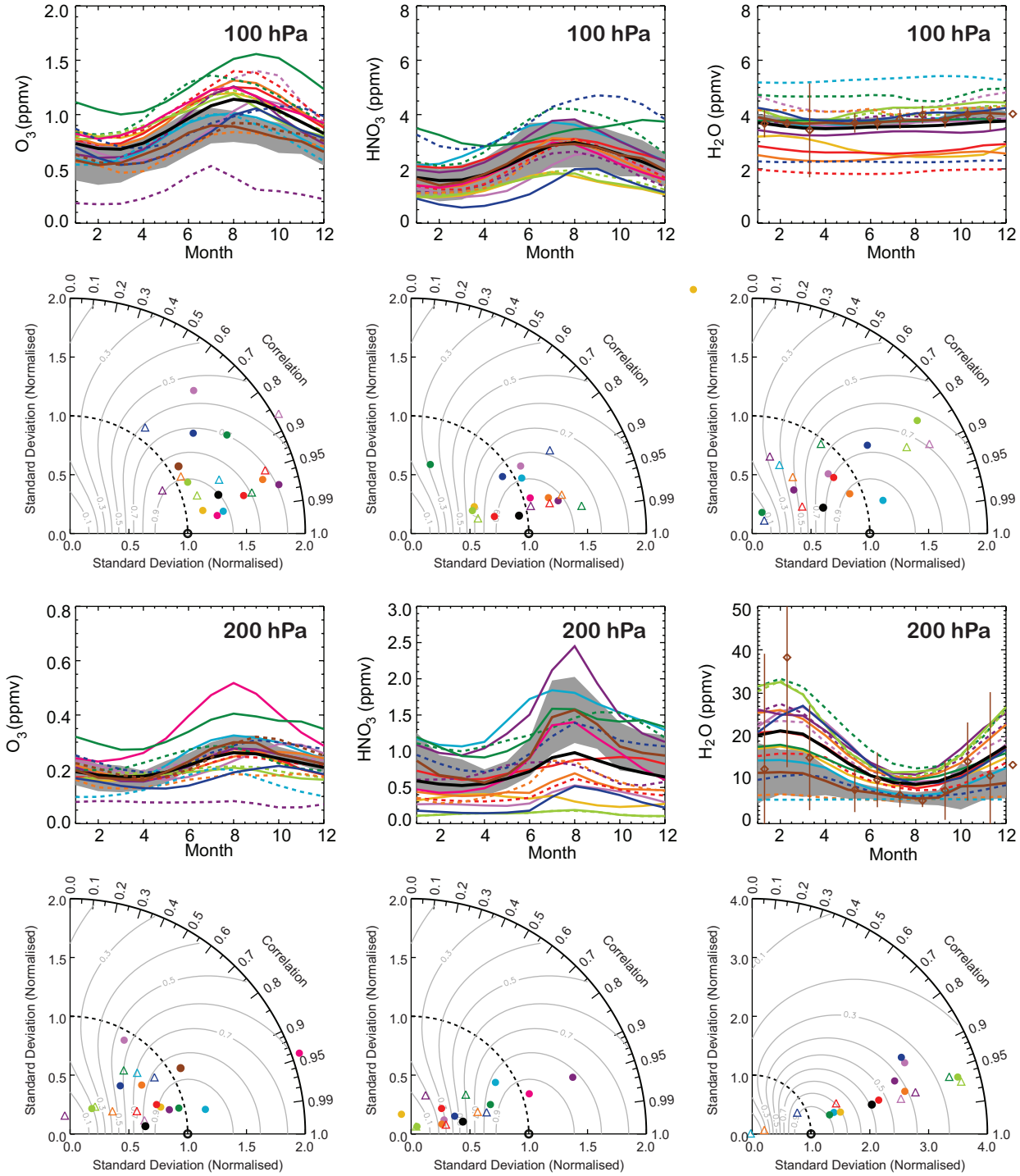


**Figure 7.23:** Same as Figure 7.22, but for latitudes between 40°S and 60°S.

The upper two rows in **Figures 7.22** and **7.23** show the results for the 100 hPa level with the corresponding Taylor diagrams in the NH and SH, respectively. In the NH, $O_3$ is relatively well represented in all the models despite a tendency to overestimate the mean and the amplitude (*i.e.*, standard deviation) of the seasonal cycle relative to MIPAS observations. Moreover, the Taylor diagram reveals slightly lower correlation values than average for CAM3.5, ULAQ, UMUKCA-UCAM, and WACCM. The seasonal cycle in $HNO_3$ mostly confirms this behaviour, with the exception of UMUKCA-METO, which exhibits a very low correlation with MIPAS satellite observations. The seasonal cycle in $H_2O$ is similar to that obtained in the tropics at 80 hPa with a several month lag in both models and observations (see Figure 7.10), pointing toward a strong connection between the tropics and the extra-tropics. The performance

of the models therefore strongly depends on their ability to represent tropical processes such as dehydration (see Section 7.4.5). Indeed, models that score low in the tropical $H_2O$ diagnostic also score low in this diagnostic. From this, it might be inferred that the slightly too high amplitude in $O_3$ has its origin in the tropics. This also confirms the finding by Gettelman *et al*. (2009) that most of the CCMVal-1 models have $O_3$ in the tropics that increases too quickly at and above the tropopause. This discrepancy is improved in the CCMVal-2 models (see section 7.4.4), but several outliers still exist. The seasonal cycles in the SH generally show smaller amplitudes, reflecting the weaker BDC, and also weaker transport within the tropically controlled transition region. The models generally show the same behaviour, however, overestimating the mean $O_3$ values. In both the NH and SH at 100 hPa, the multi-model mean reaches grades comparable to the better performing models.

The results for the 200 hPa level are shown in the two lower rows in Figures 7.22 and 7.23. At 200 hPa in the NH, the models' performance seems to decrease compared to the 100 hPa level in almost all the models. The mean values and amplitudes in the $O_3$ seasonal cycle tend to be lower than those in the observations. The worst scores are obtained by CNRM-ACM, NiwaSOCOL, SOCOL, and UMETRAC, which show too low amplitudes and relatively low correlations in comparison with the MIPAS observations. $HNO_3$ again shows a consistent behavior with that of $O_3$ in almost all of the models. The generally low mean values in both $O_3$ and $HNO_3$ can be explained by too much vertical transport across the extra-tropical tropopause. This is reflected also in too large amplitudes in the $H_2O$ seasonal cycle (as seen in the Taylor diagrams with the standard deviations on the radial axes). Tropospheric influence seems to be particularly high during late summer and autumn. As the analysis is done on fixed pressure levels, the model biases could in principle stem from biases in the tropopause altitude. However, this seems not to be the case. MRI for example shows a too low tropopause, but too strong mixing, while UMSLIMCAT shows a too high tropopause, but not enough mixing. SOCOL and NiwaSOCOL both are too

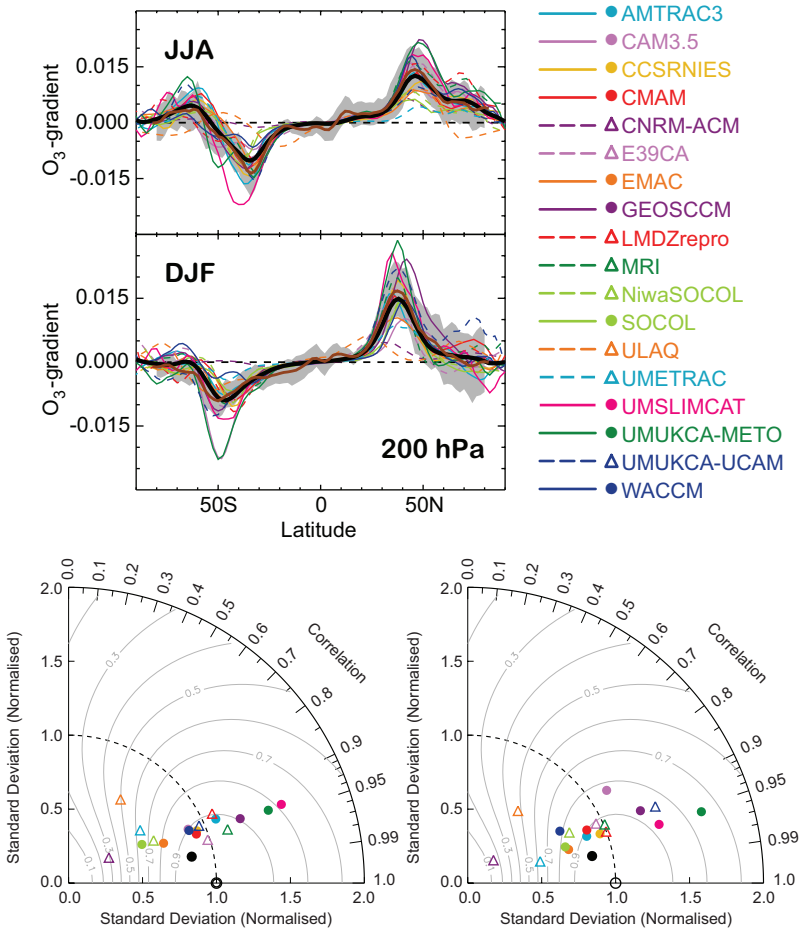


**Figure 7.24:** Meridional gradient in $O_3$ (ppmv/deg) at 200 hPa and corresponding Taylor diagrams for JJA (left panel) and DJF (right panel). Brown lines indicate MLS data averaged over the years 2004-2008, black thick lines and dots the multi-model mean. The gray region in the top panel indicates 1σ from the observations.

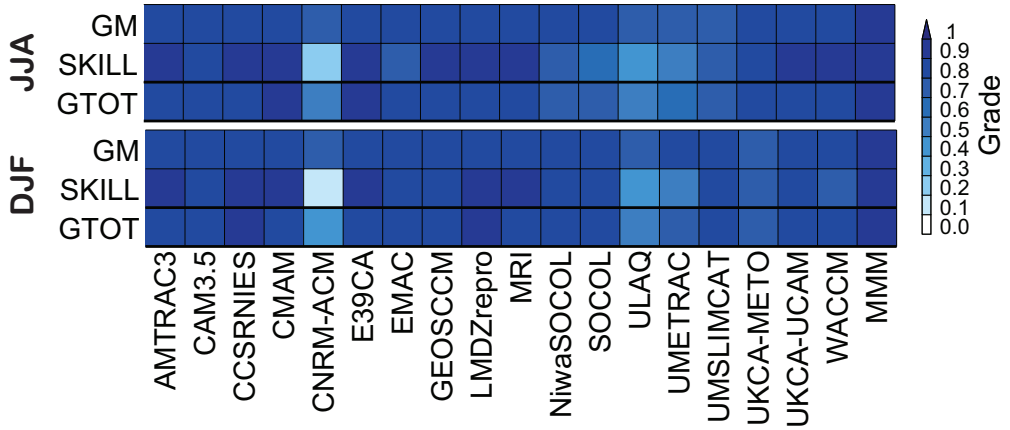


**Figure 7.25:** Same as Figure 7.17 but for meridional gradient in $O_3$ at 200 hPa for JJA (upper) and DJF (lower).

diffusive, possibly due to their semi-Lagrangian transport scheme.

At 200 hPa in the SH, the observed seasonal cycles of all three tracers show smaller amplitudes than in the NH, similar to the finding on the 100 hPa level. The models' means and amplitudes (standard deviations) in $O_3$ and $HNO_3$ are shifted to smaller values than expected from the observations. For $O_3$, the worst model performance is found for CNRM-ACM, the SOCOL-based models, ULAQ and UMSLIMCAT. Again, $H_2O$ indicates too strong cross-tropopause transport. Note that there is some evidence that the seasonal cycle in the MIPAS $H_2O$ exhibits too small amplitude. Comparison with the ACE-FTS measurements indicates that MIPAS might be somewhat low especially during summer on the 200 hPa level and in both hemispheres, an issue which is currently under investigation. However, the large noise and standard deviations in the ACE-FTS data imply that the sampling from the ACE-FTS is not sufficient in determining the seasonal cycle in $H_2O$ accurately. Additional measurements with higher (spatial and temporal) resolution will be needed to resolve this issue and to gain more confidence in this metric in the future.

For this diagnostic we derive two grades for the models. One grade is based on the $O_3$ seasonal cycle, and calculated as the average over all mean and skill values obtained for both pressure levels and hemispheres. The other grade is based on the $H_2O$ seasonal cycle, calculated as the average of all skill values obtained for both pressure levels and hemispheres. We do not include the mean for $H_2O$, since the mean is already used as metric in the tropics and we do not expect significant changes due to the $CH_4$ oxidation. The models' final grades are shown in Figure 7.39.

### 7.5.2.2 Meridional Tracer Gradients

Useful information on mixing barriers and therefore the degree of isolation and chemical distinctness of different regions such as the tropics and the extra-tropics is provided by the sharpness of meridional gradients of long-lived species. Here we use the meridional gradient in $O_3$ at 200 hPa (which is long-lived in relation to the transport time scales in this region). We use seasonal means for JJA and DJF derived from monthly mean zonal-mean $O_3$ fields (REF-B1 simulations) from all models and compare them to a multi-year seasonal climatology using MLS data (averaged over 2004-2008).

**Figure 7.24** shows that the models reproduce the meridional gradients in both seasons (JJA and DJF). Most models are within 1 sigma of the observations (gray). This implies that the models are capable of reproducing the separation between the tropical UT and the extra-tropical LMS. As can be seen in the Taylor diagrams in Figure 7.24 (lower panels) the correlations are mostly higher than 0.9, except for ULAQ and CNRM-ACM, which show correlations between 0.5 and 0.7. There is, however a substantial spread in the models in terms of standard deviations, resulting in somewhat decreased skill (see **Figure 7.25**). Too low variability is shown by ULAQ, CNRM-ACM, NiwaSOCOL and SOCOL, and too high variability by UMUKCA-METO and UMSLIMCAT, although the latter achieves a very high skill of 0.9 due to a high correlation with the observations.

A relation between this diagnostic and the zonal-mean zonal wind would be expected, as the subtropical jet acts as a barrier to transport, and maintains strong gradients across this region as observed in aircraft observations (Ray *et al*., 1999). This is indeed the case for ULAQ, which shows low grades for both the zonal-mean zonal wind and for the meridional tracer gradient.

### 7.5.2.3 Normalised Vertical Profiles of CO in Tropopause Coordinates

To evaluate the representation of tropospheric influence on the background LMS in the models, and to separate between transport across the extra-tropical tropopause

on short time scales and transport from the tropics and subtropics on longer time scales, we use CO with a ~3 month lifetime in the LMS. In the middle stratosphere above $\theta$ = 500 K, CO is nearly constant, with an observed background value of 10-15 ppbv (Flocke *et al.*, 1999), due to the chemical equilibrium between methane and CO oxidation. Any excess CO must then originate from the troposphere.

To examine the coupling between the LMS and the extra-tropical troposphere, CO was evaluated in tropopause coordinates, expressed in potential temperature units relative to the 2 PVU surface ($d\theta$) as applied to the SPURT data set (Hoor *et al.*, 2004, 2005). Key results from Hoor *et al.* (2004, 2005) are:

- The coupling to the local troposphere drops below 25% over the lowest 30 K above the tropopause (2 PVU).
- The stronger influence of the sub-tropical troposphere above the extra-tropical tropopause ($d\theta \geq$ 30 K) accounts for the background CO in the LMS, which varies with season.

The largest inter-seasonal differences are found when comparing winter/spring to summer/autumn (Hoor *et al.*, 2005).

For CCMVal-2, instantaneous model output for the year 1995 was sampled within the SPURT measurement domain (30°N-80°N, 20°W-10°E). Data were analysed in layers of 30 K relative to the 2 PVU surface (represented by the centred layer means at -15, 15, 45, and 75 K in **Figure 7.26**). The tropospheric fraction of CO in the stratosphere (CO*) is determined by CO* = (CO-$CO_{strat}$)/($CO_{trop}$-$CO_{strat}$). The stratospheric CO-background ($CO_{strat}$) was deduced for each individual model for $\theta$ = 500-600 K, for $CO_{trop}$ the layer mean for $d\theta$ = -30-0 K was used. Note, that the normalisation accounts for the varying boundary specifications of CO in the models, which would lead to a degradation of the performance in many models if not accounted for. Models that did not provide instantaneous tropospheric CO were not included in the comparison.

Two properties were tested:

1. The abundance of tropospheric tracer CO* between 30 and 60 K above the tropopause as a measure for tropospheric influence.
2. The decreased coupling to the local tropopause in the lowest 30 K above the dynamical tropopause, and at $d\theta$ = 30-90 K as represented by the different gradients of CO* in the respective layers.

For the grading, the following properties were used:

1. W1: The abundance of CO* in the $d\theta$ = 30-60 K layer was compared to the SPURT data. A model was given a grade of 3 if the difference between observations and model was smaller than 1$\sigma$ of the measured interannual variability, 2 points for $\sigma$ between 1 and 2$\sigma$, 1 point for $\sigma$ between 2 and 3, and 0 for data outside the 3$\sigma$ level.
2. W2: The vertical gradient of CO* in the $d\theta$ = 30-90 K region must be much smaller than it is closer to the tropopause (up to $d\theta$ = 30K) since the decrease in CO* is largest in the lowest layer. Grading was performed by calculating the gradients in CO* between the levels $d\theta$ = 45 and 75 K, and $d\theta$ = -15 and 15 K, respectively, taking the ratio between these two gradients, and comparing the ratios obtained from models and observations. As in W1, the ratio of the gradient of each model was tested to see if it fell within the uncertainty range of the observed ratio in steps of 1 s.

Thus, high values for both weights indicate a good separation from the extra-tropical troposphere and mainly weak influence from the subtropics (*e.g.*, CAM3.5, CMAM, WACCM, EMAC). Low values for W1 (abundance), but high values for W2 (separation) indicate too much tropospheric tracer CO* in the LMS, but the transition from the local troposphere to the LMS occurs correctly within $d\theta$ = 0-30 K (*e.g.*, AMTRAC, CCSRNIES). Low values for both W1 and W2 indicate that the coupling to the extra-tropical tropopause extends too deep into the stratosphere (*e.g.*,

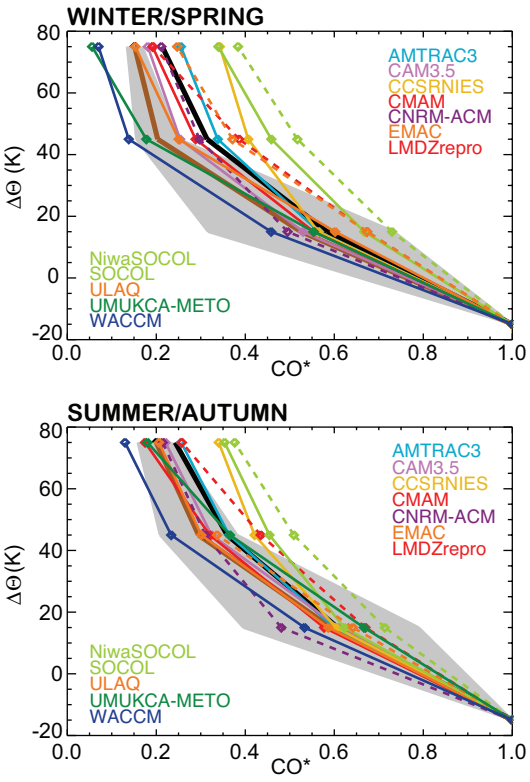


**Figure 7.26:** Profiles of CO* (normalised CO) for winter/spring and summer/autumn in layers of dθ = 30 K and for the different models. The brown solid line and grey shading shows mean CO (±1σ) from SPURT aircraft measurements. The black indicates the multi-model mean.

LMDZrepro, NiwaSOCOL during winter) leading to an unrealistic tropospheric contribution due to overestimation of transport across the extra-tropical tropopause. SOCOL, ULAQ, and CCSRNIES do perform well in both metrics during summer, but less so during winter.

In general models tend to transport too much tracer into the LMS in winter as indicated by the low values of W1, and the fact that the multi-model mean lies outside of the 1 sigma range of the observations. However, most models capture the separation (*i.e.*, the change of gradient) around $d\theta$= 30K well, as indicated by W2. Thus most models are able to separate between transport across the local tropopause in the extra-tropics and processes involving other time scales and source regions. During summer, when tropospheric influence from the subtropics is higher, the models capture this feature. The high summer values of W2 are a result of weaker differences of the vertical gradient through this enhanced transport from the subtropics accompanied with larger variability in the measurements. Most models therefore tend to get the separation between different regimes in the LMS right within the measurements' variability.

The best representation of transport and troposphere-stratosphere coupling is seen in CAM3.5, CMAM, CNRM-ACM, EMAC, UMUKCA-METO and WACCM, whereas LMDZrepro and NiwaSOCOL seem to be too diffusive or too permeable across the tropopause, confirming the results of the previous diagnostic using seasonal cycles.

### 7.5.2.4 Vertical profiles of $O_3$, $H_2O$ and CO relative to the tropopause height

The vertical structure of $O_3$, $H_2O$ and CO across the tropopause is evaluated using profiles in tropopause-referenced relative altitude coordinates (Logan, 1999; Pan *et al.*, 2004, 2007; Hegglin *et al.*, 2006; Considine *et al.*, 2008). Note, that this diagnostic uses absolute values of CO, thereby adding information on the representation of tropospheric CO to the metric of normalised vertical CO profiles discussed in Section 7.5.2.3. In the region of ±5 km around the tropopause, relative altitudes with respect to the tropopause (RALT) are effective coordinates for separating the tracer variability as a result of chemistry and transport from that caused by the variability of the tropopause height. The diagnostic requires instantaneous model output. For consistency with the coverage of the aircraft data, models are evaluated for the years between 1995 and 2005.

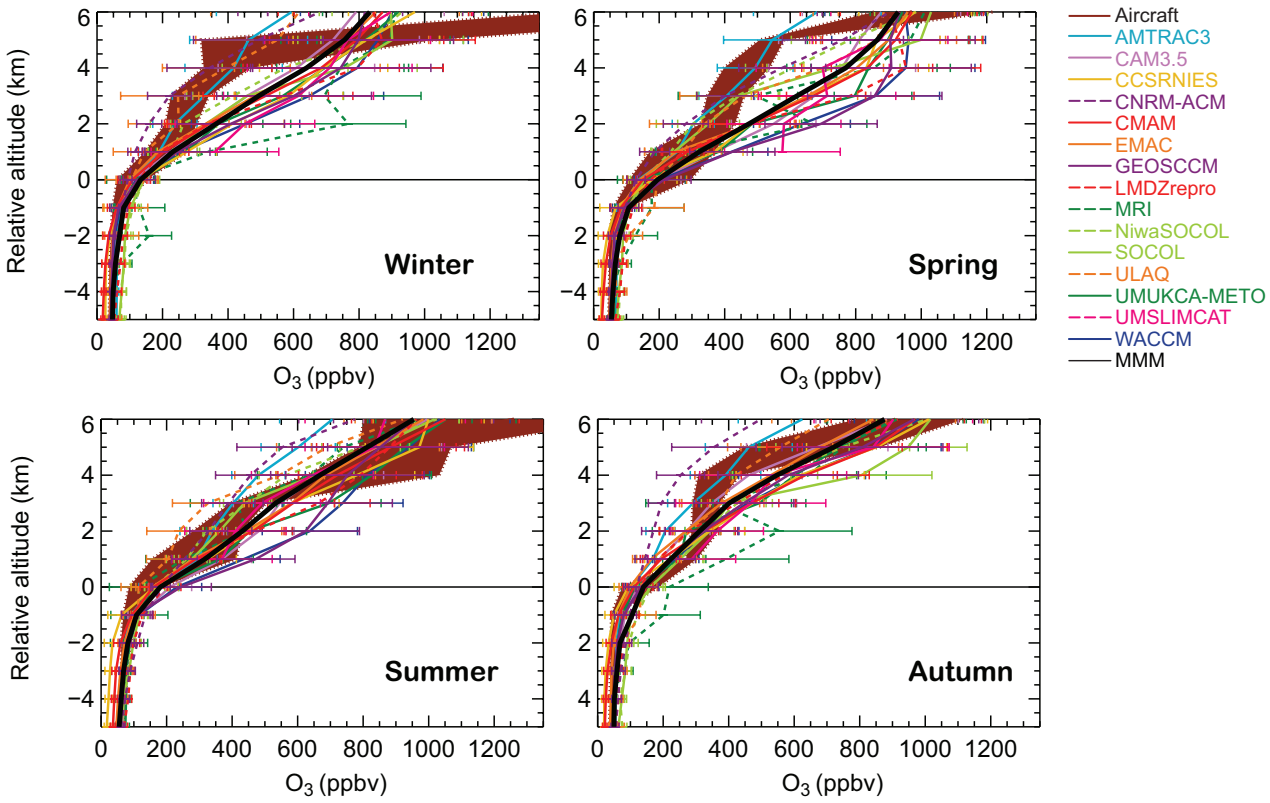


**Figure 7.27:** Ozone profiles in RALT for four seasons. The distribution of the aircraft data is represented by 25-75 percentiles (black shading) and the 5 and 95 percentiles (dotted lines). Models are represented as median (colour lines) and the 25 -75 percentiles (error bars).

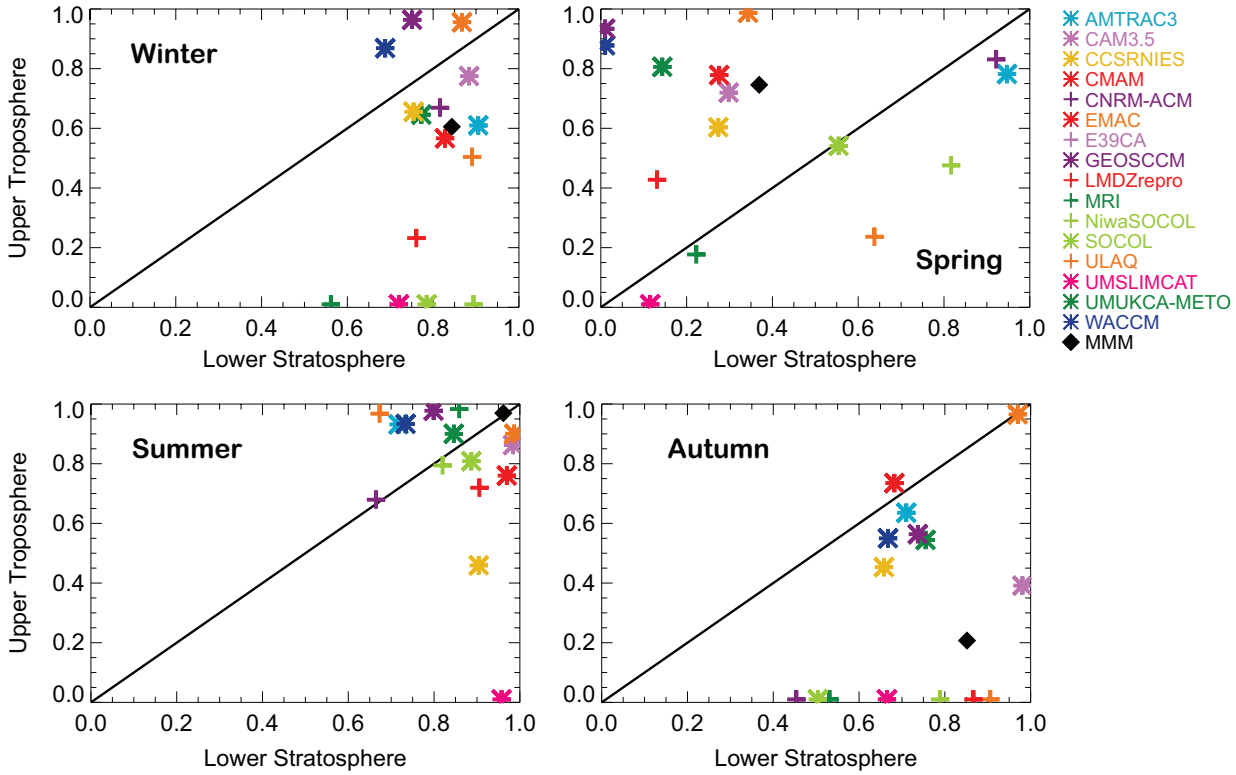


**Figure 7.28:** Grades calculated using simplified metrics (Douglass et al., 1999) for mean ozone values in the upper troposphere (1-4 km below the tropopause) and lower stratosphere (1-5 km above the tropopause) and four seasons.

The region of analysis is chosen to be that part of the extra-tropics that is not strongly influenced by the subtropical tropopause break and double tropopauses. The selection criterion is a tropopause height of 325 K or below in winter and spring, and 335 K or below in summer and fall. An additional requirement is that the latitude is lower than 80°N (Tilmes *et al*., 2010). The profiles selected are largely within 40°N-80°N. The LS chemical composition of this region, as discussed in Section 7.5, is largely controlled by the downward branch of the BDC in the stratosphere, with seasonally varying contribution from the isentropic mixing between tropical latitudes and the LMS. The region is therefore well suited for evaluating how well models represent the two competing processes. The vertical structure is examined using $O_3$, $H_2O$, and CO. The $O_3$ structure is examined for four seasons. CO and $H_2O$ structures are examined using the annual mean.

**Figure 7.27** shows the ozone profiles in relative altitudes with respect to the tropopause (RALT) for four seasons. Observations show a seasonality of ozone mixing ratios in the LMS, with lower values in fall and winter, and higher values in summer and spring. Figure 7.27 can be compared to Figure 7.22 at two pressure levels. The figures are consistent, but interpretation across coordinates is difficult because the RALT level in Figure 7.27 varies in pressure throughout the year. In general, all models represent the ozone behaviour well, qualitatively. Quantitatively, in most models ozone increases more rapidly above the tropopause compared to the aircraft climatology, a result that is also consistent with the finding of the diagnostic of the seasonal cycle in $O_3$ at 100 hPa (see Section 7.5.2.1). The comparison is quantified using the metrics defined by Douglass *et al*. (1999), as described in Section 7.3 with $n_g$ chosen to be 3. The calculated grades for the UT and LS are given in **Figure 7.28**.

In general, models agree better with observations in summer, when stratospheric transport processes are weaker compared to other seasons, and photochemical production is more active. Models also do better in the UT in spring. There is a wide spread in model performance during winter and fall.

The annual mean CO and $H_2O$ vertical structures and the grading values are shown in **Figure 7.29**. The UT chemical composition, especially CO, is significantly influenced by the contribution of anthropogenic and fire emissions in the densely populated NH. Most of the models are not representing latitudinal and seasonal variations in CO emissions, resulting in significant deviations between models and observations in UT CO distributions (Figure 7.29, left column), but good agreement is found for the LS in general. The disagreement in the UT was not identified in the previous metric (Section 7.5.2.3) due to the applied

normalisation. (The normalisation is designed on purpose to avoid testing the boundary CO condition in the troposphere, and therefore to be able to solely focus on transport and mixing effects.) $H_2O$ is well simulated by the models in both the UT and LS, except for MRI which shows too high values in the LS. The comparison with the ACE-FTS satellite data indicates a good agreement between the two data sets. They agree within their uncertainties. Differences may be due to both, the coarser vertical resolution in the satellite data and the smaller regional coverage of aircraft observations.

We use the average over the UT and LS grades of $H_2O$ for each model (shown in lower left of Figure 7.29) in Figure 7.39. We do not use the same composite grade for $O_3$ and CO, since these two species are strongly influenced by the representation of tropospheric chemistry, which is treated in most models in a simplified way.

### 7.5.2.5. Structure of the ExTL

The extra-tropical tropopause transition layer (ExTL) is composed of air masses with partly tropospheric, partly stratospheric characteristics. The representation of the ExTL characterizes how well the models reproduce the tropopause as a chemical transport boundary and the sharpness of the barrier. The transition layer depth is examined using tracer correlations between ozone and water vapour (Pan *et al*., 2007; Hegglin *et al*., 2009). A stratospheric branch is identified using a fit to a polynomial function of second order to all data points in the LS (~below 20 km) with $H_2O$ < 10 ppmv. Similarly, a tropospheric branch is represented by a linear function derived in fitting all data points with $O_3$ < 100 ppbv for both observations and models. Mixed air masses are identified as those points outside the 3 sigma range of both the stratospheric branch and the tropospheric branch. The observed transition layer is derived using POLARIS aircraft data, which include measurements in spring, summer and fall (Pan *et al*., 2007). Results of both model and observations are shown in **Figure 7.30** (left) as histograms of the fraction of samples in relative altitudes. Model analyses used output for the same seasons.

Two parameters are used to quantify the comparisons: a) the centre of the transition layer, defined as the centre point of the distribution at the half maximum, b) the width of the layer, defined as the width of distribution at the half maximum. These criteria are influenced by the bin size, which was chosen to be 0.5 km for the observations and 0.5 or 1 km for the models depending on which number is closer to their vertical resolution in the UTLS. A comparison of layer width and centre location is shown in Figure 7.30 (right). The transition layer between UT and LS is well manifested in all models, however in all cases the layer is broader, between 2 and 4 km, compared to 1 km derived from observations. Further, the layer centres in the mod-

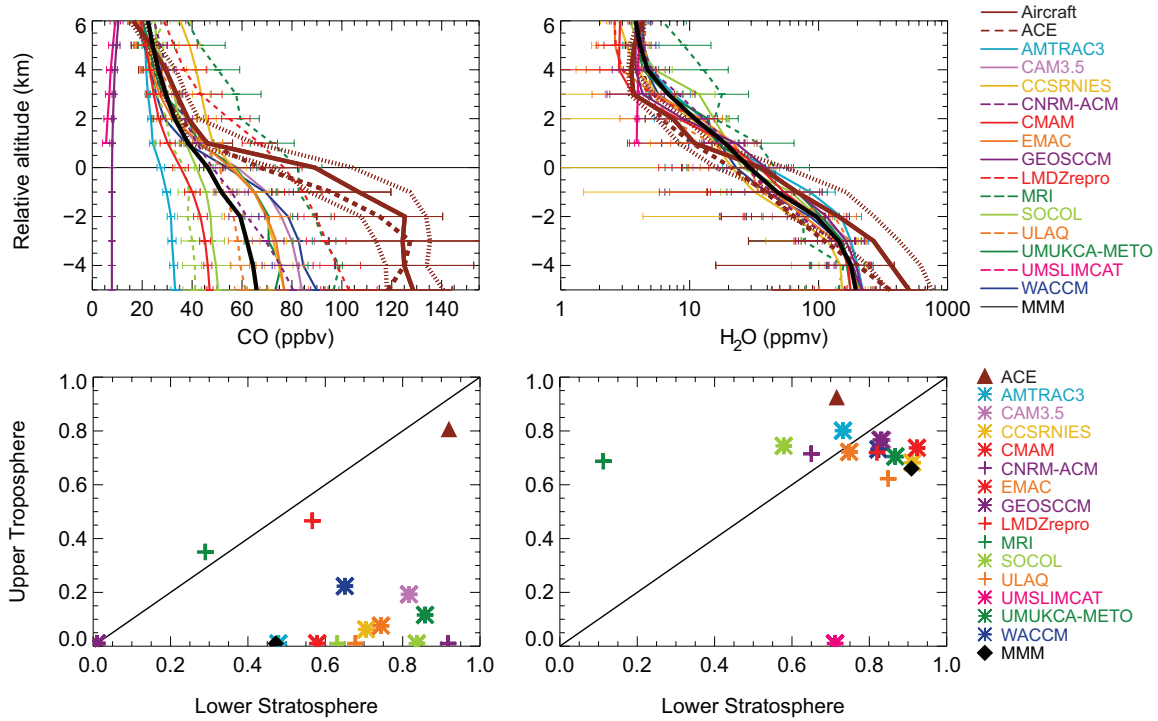


**Figure 7.29:** Top panels as in Figure 7.27 and bottom panels as in Figure 7.28, but for annual means of CO (left) and $H_2O$ (right). Also included in the comparison are the annual means of ACE-FTS data for the year 2007 (brown dashed line in the upper panels, brown triangles in the lower panels). The multi-model mean is indicated with a black solid line or diamond, respectively.

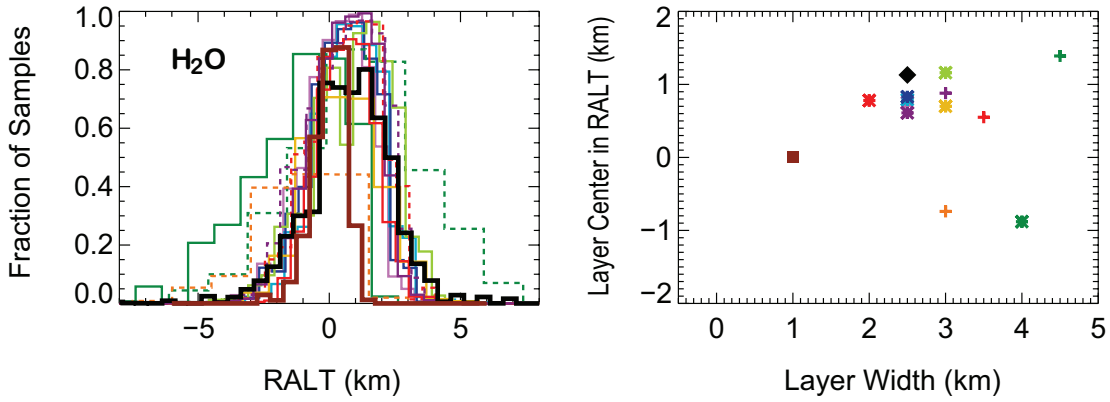


**Figure 7.30:** Left panel: Fraction of air parcels within the mixing layer from models (colours) and the multi-model mean (black) for the year 2000, and from POLARIS observations for 1997 between spring and fall (brown). Right panel: Scatter plot between centre and width of the ExTL. See legends Figure 7.27 and 7.28 for colour coding.

els are shifted upward ~1 km in most cases. Uncertainties in tropopause location derived from the relatively coarse vertical resolution models may have contributed to the discrepancies, equivalent to an artifact seen when using low resolution radiosonde observations to deduce tropopause heights (Bell and Geller, 2008).

The fact, that models have major difficulties to reproduce this diagnostic, might be at least partly attributable to the limited vertical resolution of the models (see also Section 7.5.1.4). Recent satellite observations from the ACE-FTS, which have an effective resolution similar to the CCMs (around 1 km), show indeed a behaviour similar to that of CMAM and CAM3.5, exhibiting a layer width of 2 km and a layer centre at 1 km above the thermal tropopause. The POLARIS data set on the other hand might not be representative of the ExTL depth in a climatological sense.

### 7.5.3 Variability in UTLS ozone

Correlations between tracers and tropopause heights can be used to examine the tracer sensitivity to tropopause changes in the UTLS and to evaluate transport processes between low latitudes and mid-latitudes related to synoptic waves. Here we provide a representative metric to facilitate evaluating transport and mixing processes in CCMs using conditional probability density functions (PDFs) (Rood *et al.*, 2000) of correlations between ozone and tropopause height at mid-latitudes. Simulations are compared to version (v2.2) of MLS observations between 2004 and 2008.

Whenever tropopause relative altitude is used as the vertical coordinate, it should be kept in mind that a difference in vertical profiles might simply result from time-varying tropopause height levels and consequent changes in reference levels (*i.e.*, zero levels) in the coordinate.

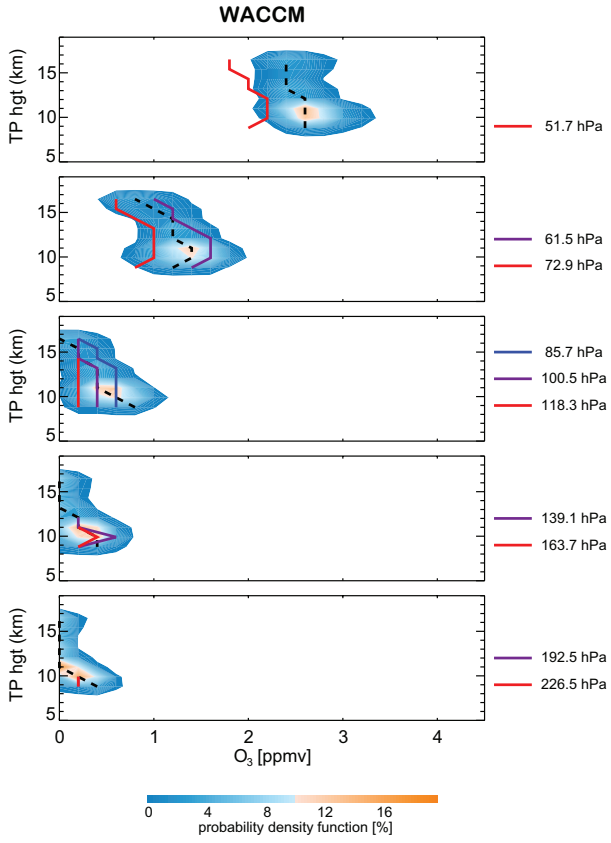


**Figure 7.31:** Example of probability function maps at NH mid-latitudes (40°N to 50°N) during JJA season showing the relationship of $O_3$ (x-axis) and tropopause heights (y-axis) for MLS (shaded contour) at five isobaric levels (46.4, 68.1, 100, 146.8 and 215.4 hPa, from top to bottom) and WACCM (coloured lines). Lines indicate the maxima in the MLS (dashed) and model (solid) PDFs at each tropopause height for each pressure level.

In other words, even the same levels in the tropopause-relative-altitude coordinate may contain tracers at different levels of pressure. Our key concern is whether there would be changes in tracer distributions due to the varying tropopause height that would distort the interpretation of changes in distributions due to transport and mixing processes by synoptic waves. To examine the tracer sensitivity of models to the tropopause changes at mid-latitudes, two dimensional correlation maps between simulated tropopause heights and ozone ($O_3$) from REF-B0 CCM simulations at five isobaric surfaces (46.4, 68.1, 100, 146.8 and 215.4 hPa) in the UTLS region are compared to maps with MLS observations. An example is shown for WACCM in **Figure 7.31**.

At NH mid-latitudes there are distinguishable changes in the observed vertical profiles of $O_3$ by varying tropopause heights during the summer. For higher tropopauses, $O_3$ exhibits lower values throughout the UTLS. The decreased $O_3$ related to increased tropopause heights is largest at 68.1 hPa. When the tropopause height is above 15 km, 68.1 hPa $O_3$ is 50% lower than $O_3$ with a tropopause height below 10 km. Therefore, when tropopause heights are higher, decreased $O_3$ can be commonly found in both MLS and CCMs.

The PDF representation of model/data comparison is entirely qualitative. To provide a quantitative measure of the performance of the CCMs, the simple metric *g* defined in Douglass *et al*. (1999), which is similar to Equation (7.1), is applied to this diagnostic. Here we used $n_g = 5$, thereby expanding the range of acceptable performance (larger difference from observed for a given *g* value). To calculate *g*, simulated $O_3$ concentrations on the 68 hPa surface were estimated using linear interpolation. Then for each month, we classify all observed and simulated data within a latitude band 40ºN-50ºN into two groups, for high and low tropopause cases. A tropopause height of > 14 km was used to distinguish tropical air masses from air originating at mid-latitudes. In each group, the average and standard deviation of $O_3$ are calculated. Comparing the difference in each group between MLS and models allows evaluation of models relative to observed data, minimising effects due to different frequency of higher tropopause heights between the real atmosphere and the CCMs. Also the difference in $O_3$ concentrations between the two groups shows the dependence of tracer concentrations on tropopause heights.

**Figure 7.32** displays the grades (*g*-values) derived from the differences between the maximum in the $O_3$ PDFs of the models and MLS. In winter, most models show good agreement with MLS except UMUKCA-METO, which shows much higher $O_3$ values (not shown). Overall, models do worse in the summer than winter. The common weakness of the other five models is for high tropopause cases during the summer-time. When the tropopause height is lower than 14 km, models show better agreement with

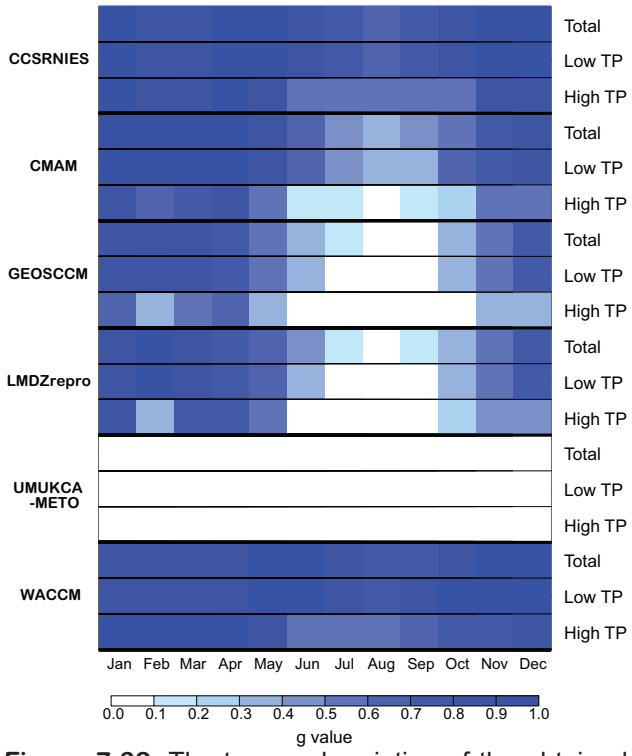


**Figure 7.32:** The temporal variation of the obtained grades (g values) on 68 hPa isobaric surfaces at mid-latitude (40ºN – 50ºN). From top to bottom, g values (calculated after Douglass et al., 1999) for CCSRNIES, CMAM, GEOSCCM, LMZDrepro, UMUKCA-METO and WACCM. In each model, g values for all tropopauses, low tropopauses (≤ 14 km) and high tropopauses (> 14 km) are shown separately.

MLS. Therefore, to improve performance of models in the summer, models need to better reproduce $O_3$ when the tropopause is higher.

To examine the dependence of $O_3$ on tropopause height, we calculated mean $O_3$ for low and high tropopause cases separately and subtracted $O_3$ for high tropopause cases from low tropopause cases. The resulting $O_3$ difference between the high and low tropopause cases has an annual cycle both in the MLS data and the models (**Figure 7.33**). This annual cycle may be related to stronger wave activity and accompanying mixing in the winter and spring. Simulated $O_3$ is less sensitive to the changes of tropopause heights than in MLS, with smaller simulated $O_3$ differences between low and high tropopause cases than observed. The smaller differences simulated in the models are likely the main reason for discrepancies shown in Figure 7.33 and lower overall performance of models in the summer. This is likely a result of more frequent high tropopause cases at mid-latitudes during summer than in other seasons, which is possibly due to a connection with stronger moist-convective events in this season, which is likely under-estimated by the models.

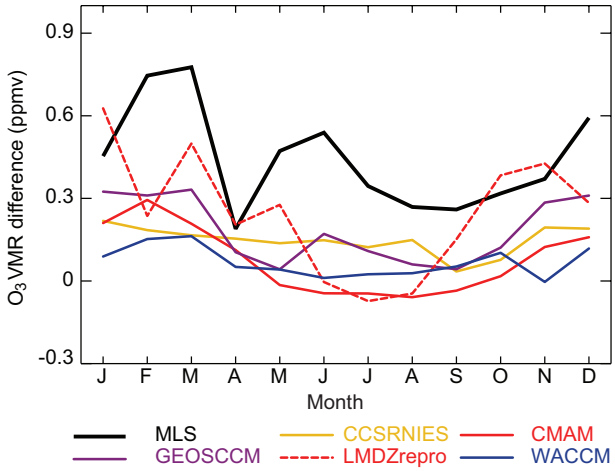


**Figure 7.33:** Time series of $O_3$ differences ($O_3$ for low tropopause – $O_3$ for high tropopause) on 68 hPa isobaric surfaces from MLS observation (black) and five model outputs. The average of $O_3$ differences is taken at NH mid-latitude (40ºN – 50ºN) and is calculated each month.

Conventional inter-comparison tools that simply analyse the difference of values between models and observations do not illuminate model performance related to the synoptic wave activity and accompanying changes in tropopause heights. The sensitivity of $O_3$ at isobaric levels highlights the importance of tropopause heights to simulated tracer distributions at mid-latitudes especially in the summer. Tropopause heights could affect average and variation of tracer fields in the UTLS. Note however that the performance of models is correlated tropopause height variability, not with mean tropopause height.

## 7.6 Trends in the UTLS

The CCMVal model runs for the past and future provide a unique multi-model ensemble to examine trends in the UTLS. UTLS trends for CCMVal-1 models, and for IPCC AR4 models, have recently been analysed by Gettelman *et al*. (2009), Son *et al*. (2009a) and Son *et al*. (2009b). Trends have also been presented for REF-B1 historical simulations in the context of validating the models against observations. Here we present some basic results of future trends in the UTLS from CCMVal-2 models.

### 7.6.1 Tropical Tropopause Trends

We have already shown historical trends for cold point temperature (Figure 7.4) and tropopause pressure (Figure 7.6), along with available data from reanalysis systems and radiosonde observations. These will be discussed further here with the addition of trends for the future (REF-B2) runs. We note that for convenience, future runs were processed using zonal-mean data. As noted by Son *et al*. (2009a) 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 and calculating tropopause pressure and cold point temperature 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.

Tropopause pressure in the models over the historical period is well constrained as noted above. The robustness of the tropopause pressure metric 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. Analyses have trends of -0.4 hPa/decade, and models are slightly higher (-0.3 to -0.9 hPa/decade). The four models with the highest overall on all 4 metrics have a mean trend of -0.6 hPa/decade. Interannual variability is highly correlated with observations, and generally small. Model absolute values of pressure vary, with many close to the observations, and several typically a level above or below. Recall that all analyses and models have been processed onto a relatively high (10-15hPa) resolution vertical grid. There are generally larger decreases in pressure in the subtropics where tropopause gradients are large.

Trends for the future REF-B2 runs are illustrated in **Figure 7.34**. The figure contains multiple ensembles for WACCM and CMAM, illustrating that the trends are quantitatively the same for different ensemble members. There are some large differences in trends in the models. CMAM, UMSLIMCAT, UMUKCA-METO and CNRM seem to have trends that are larger (-10 to -15 hPa per century) than other models (-5 hPa/century). Some of the difference in the trends may be due to the resolution of the models in the TTL. The multi-model mean is about -7 hPa per century. This does not appear directly correlated with stratospheric metrics, such as changes in the strength of the BDC. Note that the TCPT and PTP mechanisms are different, as discussed in Section 7.4.2.

Tropical cold point temperature trends are illustrated for the REF-B1 case in Figure 7.6. Models do not show the cooling over the last 25 years seen in NCEP and NCEP2. However, an analysis of the distribution of the 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. The overall picture is one of cooling in some regions balancing warming, for little net trend. This indicates that TCPT patterns respond to chang-

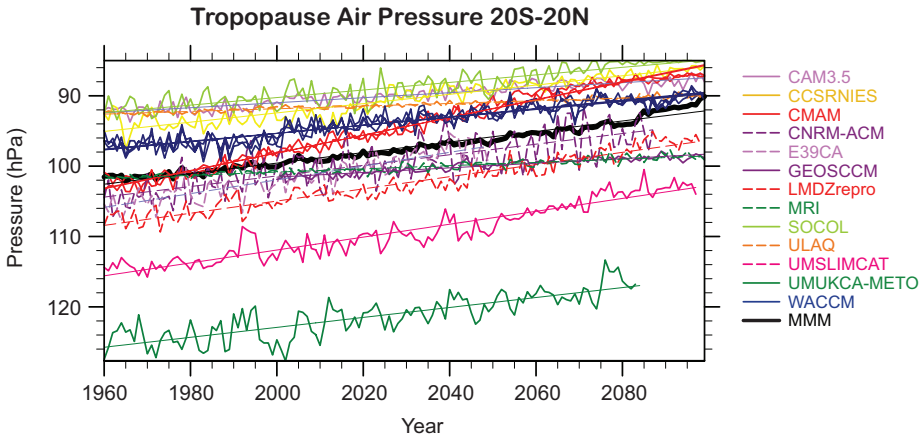


**Figure 7.34:** Lapse Rate Tropopause Pressure time series from 20°S-20°N for future REF-B2 scenarios. Thin lines are linear fits. Multi-model mean (MMM) is the thick black line.

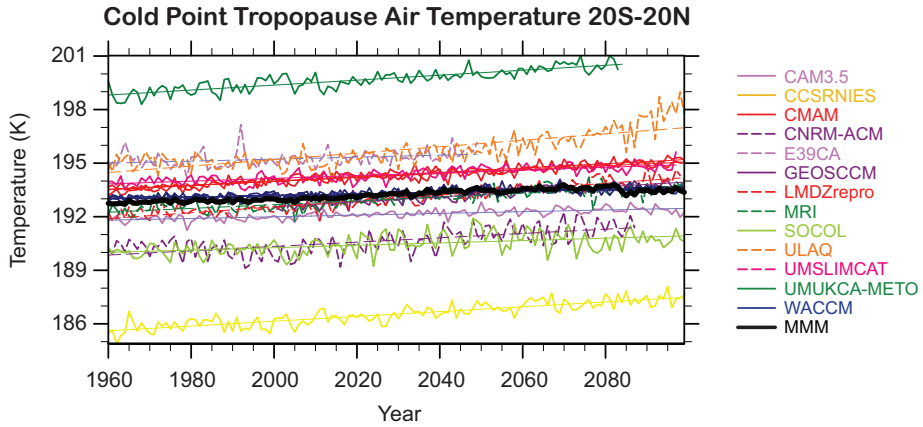


**Figure 7.35:** Cold Point Temperature time series from 20°S-20°N for future REF-B2 scenarios. Thin lines are linear fits. Multi-model mean (MMM) is the thick black line.

es 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.

Tropical cold point temperature trends for the future (REF-B2) are illustrated in **Figure 7.35**. Most models show a slow increase in minimum temperature of 0.5-1.0 K per century. Several models (ULAQ, UMUKCA-METO) have larger trends. This is consistent with water vapour increases (see below).

There exist no consistent observations of water vapour trends over long periods of time. There are indications of long-term increases in water vapour from a variety of records (Kley *et al*., 2000), and a significant increase in water vapour in the 1990s observed by HALOE, followed by a sharp, step-change decrease after 2000. The overall 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 -0.03 ppmv $yr^{-1}$ (but significant only at the 95% level). Recent changes in water vapour over the last decade or so are broadly consistent with changes in the tropical tropopause temperature (see Section 7.4.5 and Randel *et al*., 2006).

Future changes in water vapour just above the CPT simulated in the models are illustrated in **Figure 7.36**. Also illustrated in Figure 7.36 are multiple ensembles from WACCM (3) and CMAM (2), confirming that their trends are different from each other, but consistent across the same model ensemble members. Most model trends are from 0.5-1.0 ppmv per century, or nearly 25%. These trends are affected very little by methane oxidation at 80hPa, so that is unlikely to be a cause of these trends. This is consistent with the magnitude of TCPT trends, and temperature trends of 0.5-1 K per century at 193 K translate into 0.5-1 ppmv per century increase in water vapour. Models with larger temperature trends, or a stronger correlation between water vapour and temperature, appear to indicate larger increases in water vapour. This is true for example of ULAQ (large *T* increase) and MRI, CNRM-ACM and CCSRNIES (strong dependence of $H_2O$ on *T*). UMUKCA-METO is off-scale in Figure 7.11 (no correlation between $H_2O$ and *T*) so it is not surprising there is no increase in water vapour. SOCOL

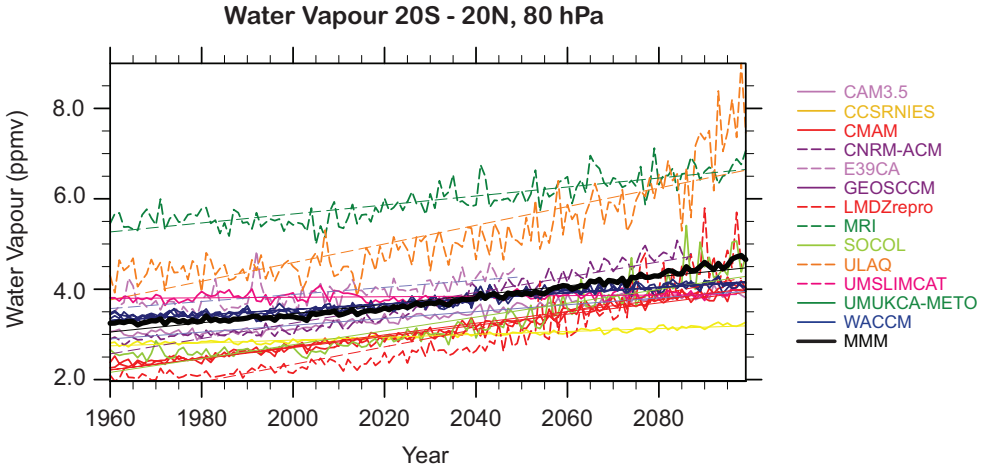


**Figure 7.36:** 80 hPa water vapour time series from 20°S-20°N for future REF-B2 scenarios. Thin lines are linear fits. Multi-model mean (MMM) is the thick black line.

indicates a large change in water vapour, without a large change in temperature. Note that the UMUKCA models (fixed water vapour) and GEOSCCM (output problem with water vapour) are not included in the analysis of REF-B2. Water vapour trends are also illustrated in **Figure 7.37**, indicating larger water vapour trends in the upper tropical troposphere at the convective outflow level near 200 hPa.

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. Generally, one expects the trends in absolute (*e.g.*, pressure) coordinates to be affected by tropopause height trends. We will therefore show

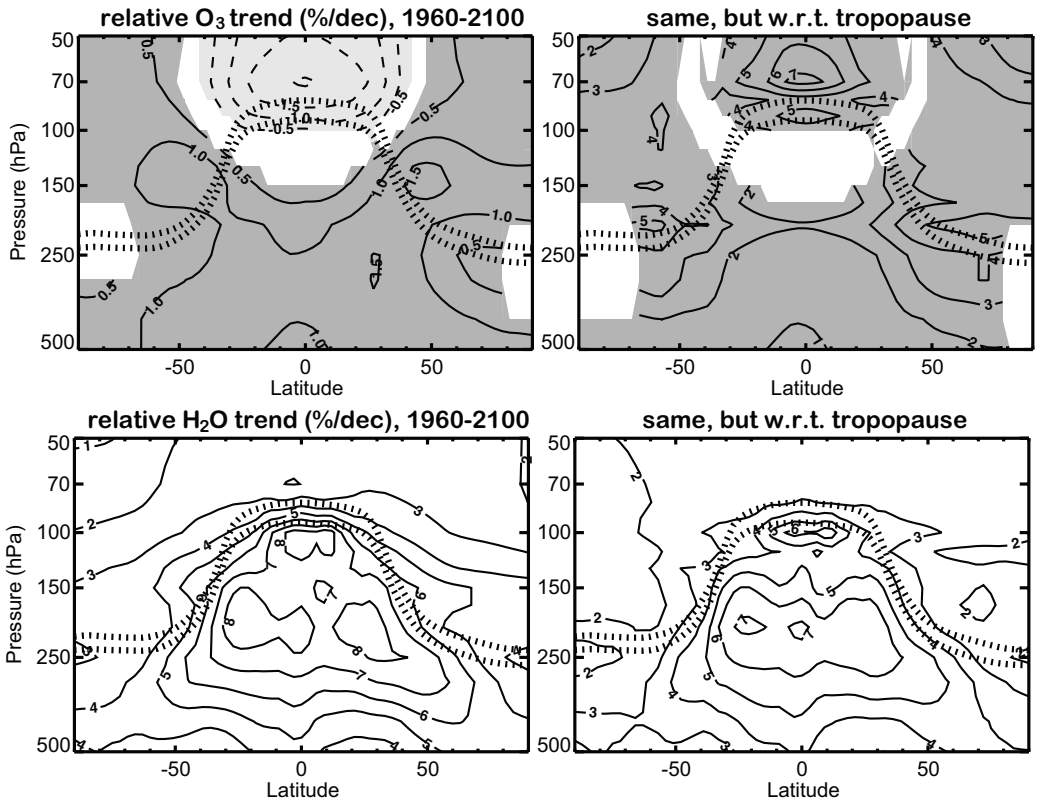


**Figure 7.37:** Trends in $O_3$ (upper panels) and $H_2O$ (lower panels) in pressure (left panels) and tropopause coordinates (right panels). Shading indicates the 95% significance level. For $H_2O$, the calculated trends are significant at the 95% level. Dotted lines in each panel denote the tropopause with the lower line corresponding to the reference period (1960-1980) and the upper line corresponding to the year 2100.

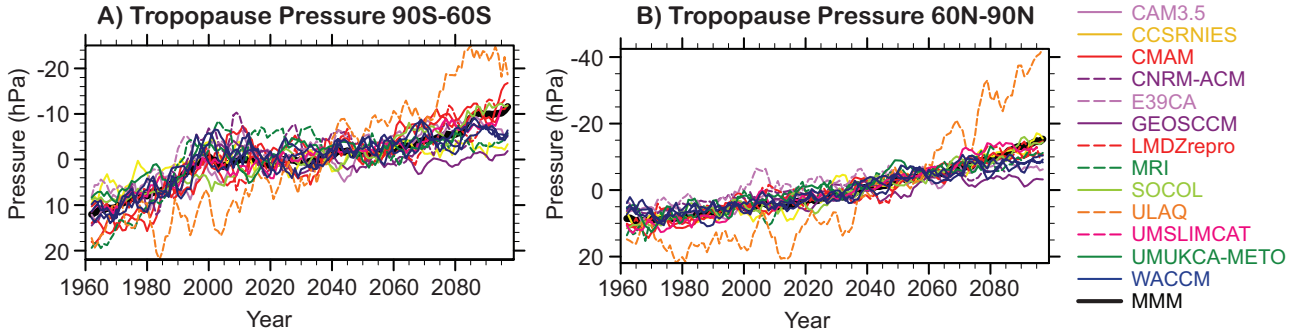


**Figure 7.38:** Northern and Southern Hemisphere extra-tropical tropopause pressure time series from 90°S-60°S (left panel) and 60°N-90°N (right panel) for future REF-B2 scenarios. Multi-model mean (MMM) is the thick black line.

two sets of trends, in absolute coordinates as well as in tropopause-based coordinates to highlight the sensitivity of trends to the tropopause. Trends are calculated as above, based on the monthly zonal mean output with respect to the tropopause obtained from the T2Mz data.

Figure 7.37 shows the 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, LMDZrepro, MRI, SOCOL, ULAQ, UMSLIMCAT, and WACCM. The left panels show trends in conventional (absolute) coordinates, whereas the right panels show 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 panels (with the higher altitude tropopause corresponding to 2100). Here, the average over the period 1960-1980 is used as reference state.

$O_3$ trends are negative (-2% dec$^{-1}$) in conventional coordinates in the tropical lower stratosphere. Decreasing $O_3$ is consistent with a strengthening of the BDC. Moderate increases of around 0.5-1.5% dec$^{-1}$ 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 trends are strongly positive above the tropopause in both the tropics and extra-tropics (4-5% dec$^{-1}$). In the tropics the sign is reversed between conventional and tropopause based coordinates. Ozone decreases due to faster upwelling resulting from an enhanced BDC. Thus $O_3$ decreases at any given pressure level. This may be a direct result of higher tropical SST (Deckert *et al*., 2008). However, the gradient of ozone around the tropopause increases as the tropopause moves to higher altitudes, so relative to the tropopause, $O_3$ increases. This 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 trends in the upper troposphere from a realistic upper troposphere (UT) base state. The base state has high humidity in tropical convective outflow regions and low humidity in downwelling branches of the Hadley and Walker circulations. In the tropical UT maximum trends of ~9% dec$^{-1}$ are found around 200 hPa. These trends are likely due to increases in temperature associated with anthropogenic greenhouse gas induced warming. In conventional coordinates one also finds rather strong positive changes throughout the extra-tropical LMS (between 3-5% dec$^{-1}$). However, these changes in the LMS are in part caused by the 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% dec$^{-1}$ throughout the global lower stratosphere.

## 7.6.2 Extra-tropical Tropopause Trends

Trends in the extra-tropical tropopause pressure for future scenarios are shown over the southern (**Figure 7.38**, left panel) and northern (in Figure 7.38, right panel) polar caps for REF-B2 simulations from 1960-2100. As in the tropics, tropopause pressure is expected to decrease -20 hPa per century in both hemispheres. The century scale trends are not quantitatively different between hemispheres over the 21$^{st}$ century (about 30hPa for the century). However, it is clear that there are differences in the timing of PTP trends between the hemispheres in Figure 7.38: 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*. (2009) in comparing IPCC AR5 models with and without ozone depletion, these differences are due to the effects of ozone depletion (1960-2000) and recovery (2000-2050). Note that the overall trend over 140 years is nearly the same in both hemispheres.

## 7.7 Summary and Conclusions

### 7.7.1 Quantitative metrics

**Figure 7.39** shows the grading obtained for the key diagnostics presented in this chapter, and provides an overall assessment of how well the models performed in the UTLS. The upper panel depicts the grades for the tropical diagnostics, the lower panel grades for the extra-tropical diagnostics as discussed in the specific sections of this chapter, and in a more qualitative way in Section 7.7.2. The grading methodology does not necessarily define an 'acceptable' grade. The different methods may yield different scores. The discussion defines what is an acceptable grade based on the uncertainty in the observations, often using the spread of multiple observation data sets. For some metrics this spread is narrow (*e.g.*, U@200 in Figure 7.17, or PTP in Figure 7.6) while for others there is large uncertainty (*e.g.*, TCPT in Figure 7.6). In general tropical grades > 0.5 are considered acceptable, and grades > 0.6 in the extra-tropics are considered acceptable. This difference might reflect the larger database of observations in the extra-tropics, or simply the choice of diagnostics.

### 7.7.2 Qualitative Diagnostics Discussion

*Tropical Diagnostics*

**Tropical Cold Point Temperatures:** The annual cycle of tropical cold point temperatures is well reproduced by most models, as is the amplitude of the annual cycle. There remain some significant biases between models, with the UMUKCA models having warm temperatures, and CNRM-ACM and CCSRNIES having cold temperatures. CNRM-ACM has too large a response to volcanic perturbations, and SOCOL and NiwaSOCOL are also high in this regard. Most models do not have strong trends in CPT over the historical period. Analysis systems also disagree over the satellite period.

**Tropical Tropopause Pressure:** Again, most models get the absolute value of tropical tropopause pressure to the right level (about 100 hPa). The UMUKCA models are below this (120hPa), which is likely the reason for their tropopause temperature warm bias. This may be a function of a slightly different vertical structure in the tropopause region. CNRM-ACM and CCSRNIES are slightly higher, as are the SOCOL models, ULAQ and EMAC. Most models do get consistent trends in tropopause pressure. Again, CNRM-ACM has too large a response to volcanic events. In general model variance is higher than observed interannual variance of tropopause pressure. Trends are consistent between models and analysis systems and variability is highly correlated.

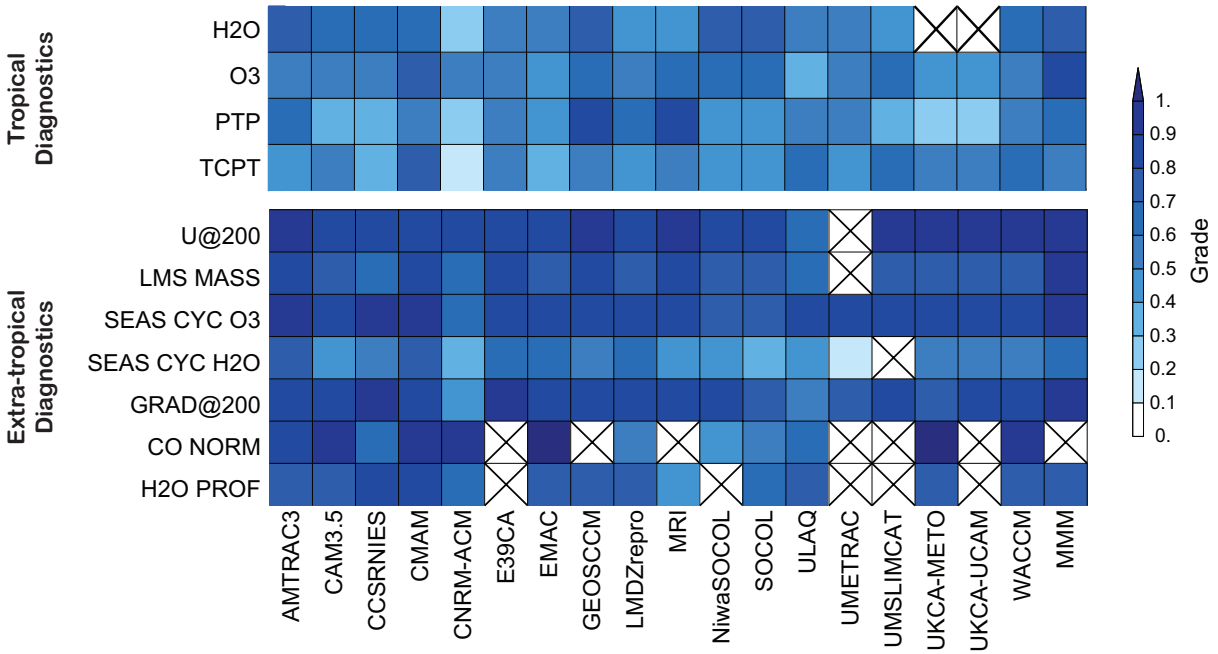


**Figure 7.39:** Quantitative metrics summary for tropics (upper) and extra-tropics (lower). MMM indicates the multi-model mean.

**Tropical Ozone:** 100hPa ozone is generally well reproduced. Most models have less NH summer-time tropical ozone than observations at 100hPa. CNRM-ACM and UMSLIMCAT have the wrong annual cycle. The UMUKCA models are higher throughout the year, as is ULAQ.

**Tropical Water Vapour:** E39CA, CNRM-ACM and MRI are too wet at 80 hPa, and several models (LMDZrepro and EMAC) are too dry, with water vapour below 3 ppmv throughout the year. The annual cycle is not as well reproduced, with many models shifted 1-2 months (or more) early relative to HALOE observations. With respect to the CPT and water vapour correlation, there are 3 models (CCSRNIES, CNRM-ACM and UMETRAC) that are clear outliers: There appears to be more water vapour than the temperatures would permit if transport were occurring similarly to observations. UMUKCA-METO is off-scale due to a very warm CPT.

**Equatorial Wave Activity:** There are huge differences between the analysis systems for the wave diagnostics, and the models examined (CCSRNIES, MRI, CMAM, WACCM) span this range. CCSRNIES appears to have lower wave activity than any of the analysis systems, which are on the low end of observed. Several models have higher ISO activity than the analysis systems (WACCM, CMAM), which is likely similar to observations.

*Extra-tropical Diagnostics*

**Seasonal zonal-mean zonal wind:** Most models perform well in this diagnostic as well, except ULAQ. This might be due to an insufficient horizontal and vertical resolution in this model and its geostrophic dynamical core.

**Seasonal cycle in LMS mass:** Most models represent well the phase and amplitude of the seasonal cycle in LMS mass, but this is not so for the annual mean value in LMS mass. Overall scores are generally higher for the NH than for the SH. Models scoring high are AMTRAC3, CMAM, E39CA, and GEOSCCM. Models which perform the least well are CCSRNIES, CNRM-ACM, and ULAQ. The diagnostic yields insight into the strength and seasonality of the BDC which will affect stratosphere-troposphere exchange and therefore both UT and LS tracer distributions.

**Seasonal cycles in $O_3$, $HNO_3$, and $H_2O$ at 100 and 200 hPa:** Most models perform reasonably well for $O_3$ in the NH, however the amplitude is generally a little too high at 100 hPa and too low at 200 hPa. The latter finding indicates that the models exhibit too much transport across the extra-tropical tropopause (in particular CNRM-ACM and the SOCOL-based models). The spread in skill in representing $H_2O$ is larger than that in $O_3$, with models doing better at 100 hPa than at 200 hPa. At 200 hPa, strong tropospheric influence causes too large amplitudes. Note, that there exist some observational uncertainties in $H_2O$ at the 200 hPa level. Better measurements are needed to gain more confidence in this diagnostic. The spread in skill to represent $HNO_3$ in the SH at 200 hPa is even larger. UMSLIMCAT shows the best representation of $HNO_3$, but most other models have a low correlation with observations and under-estimate the annual cycle amplitude.

**Sharpness of meridional gradients in $O_3$:** The results of this metric are largely consistent with the one from the seasonal cycle in $O_3$ on 200 hPa. Models perform generally well, with some models overestimating (UMSLIMCAT and UMUKCA-METO), and some models (CNRM-ACM and ULAQ) under-estimating the maximum in the gradient.

**Normalised vertical profiles of CO in potential temperature relative to the tropopause height:** Most models perform reasonably well in this diagnostic and are able to separate between transport across the local tropopause in the extra-tropics and processes involving other time scales and source regions, except LMDZrepro, SOCOL, and NiwaSOCOL.

**Vertical profiles of $H_2O$, CO, and $O_3$ in tropopause coordinates:** The models show some difficulties simulating the seasonal mean vertical profiles of the different tracers. Models perform best for $H_2O$, possibly because it is least affected by chemistry. CO is represented very poorly, clearly because most models do not include tropospheric chemistry. However, even models including tropospheric chemistry perform rather poorly in this diagnostic (CAM3.5, EMAC, and ULAQ). The lack of a more sophisticated tropospheric chemistry also causes poor UT $O_3$ distributions.

**Depth of the extra-tropical tropopause transition layer (ExTL):** The models simulated an ExTL that is deeper than observed, and shifted above the thermal tropopause. This might be due to the models' limited vertical resolution as indicated also by the findings of the study by Bell and Geller (2008). CMAM's representation of the ExTL is closest to the observations, which is noteworthy since CMAM has a relatively low horizontal resolution compared to other models. The ratio between vertical and horizontal resolution might matter more. Models that show the most difficulties in reproducing the ExTL are SOCOL, UMUKCA-METO, and CNRM-ACM, which might be due to too diffusive transport schemes. For the ExTL (as well as for the TIL), it seems difficult to find a model mea-

sure that can be usefully compared against observations due to the mismatch between the scale of observations and the scale resolved by models.

### 7.7.3 Qualitative Model Discussion

***Summary:*** In the tropics there are 4 models that score at least 0.5 on all metrics and have consistent transport and trends: CMAM, GEOSCCM, E39CA and WACCM. In the extra-tropics, models that score consistently higher than 0.6 are AMTRAC3, CMAM, EMAC, E39C and UMSLIMCAT. Note however, that the latter two are validated only on a subset of the diagnostics.

**AMTRAC3** performs acceptably well on the tropical metrics examined, although slightly lower on TCPT. Summer tropical ozone is a bit low, as with other models. Water vapour annual cycle is shifted a month or two early. In the extra-tropics, AMTRAC3 is one of the better performing models, with scores mostly equal or higher than 0.6.

**CAM3.5** shows generally good performance on tropical metrics. The tropical tropopause is slightly high and cold, but the tropical water vapour annual cycle is one of the best despite this. CAM3.5 also performs relatively well on extra-tropical metrics, with the exception of the metrics for the seasonal cycle in $H_2O$, which shows a too large (small) amplitude at 200 (100) hPa, respectively.

**CCSRNIES** has a tropical tropopause pressure is high and tropopause is cold, yet water vapour is reasonable, implying a different transport than observed. CCSRNIES scores lower on some of the tropical metrics as a result. Tropical 100 hPa ozone is relatively good. In the extra-tropics, CCSRNIES performs relatively well, with the exception of exhibiting a too small mass in both the NH and the SH LMS, and a strong shift in the seasonal cycle of $H_2O$ at 100 hPa. The normalised vertical profiles of CO indicate a good separation from the extra-tropical troposphere, but too large transport from the tropical tropopause.

**CMAM:** Tropical tropopause temperature is well reproduced. CMAM has all tropical metrics over 0.5. However, there are very large trends in tropical tropopause pressure for the REF-B2 simulations. In the extra-tropics, CMAM is one of the best performing models, with scores mostly above 0.6, and it also shows the best representation of the ExTL. CMAM performs less well for $HNO_3$ in the SH.

**CNRM-ACM** exhibits some significant problems with tropical transport. The tropopause is cold and high, with more water vapour than would be implied by the temperatures. The tropopause response to volcanic events is very large, and the secular trend is also larger than other models due to problems with volcanic aerosol heating. Future trends are also relatively large in water vapour compared to other models, and not consistent with temperatures. Also in the extra-tropics, CNRM-ACM shows major deficiencies in both the dynamical and the transport and mixing diagnostics. Extra-tropical tropopause pressure is too low, which is also reflected in too small values for the LMS mass in both hemispheres. These deficiencies go along with too low $HNO_3$ in the SH, and too high $H_2O$ in the NH at 200 hPa.

**E39CA** performs well on tropical diagnostics with a score of 0.5 on all grades. The tropical tropopause pressure variability is larger than observed, and tropical ozone is lower than observed. In the extra-tropics, E39CA in one of the best scoring models, however some critical diagnostics are missing. E39CA got only one low score of 0.5 for the amplitude in the seasonal cycle of $H_2O$ in the NH at 200 hPa, which is too large.

**EMAC:** The tropical tropopause is colder and higher, resulting in less water vapour in the stratosphere than observed. Transport appears to be correct, though without much of an annual cycle. The seasonal cycle in ozone at 100 hPa is relatively well represented. In the extra-tropics EMAC performs relatively well, especially for the metrics based on the mean $O_3$ values in the UT and LS, the metric for CO in $d\theta/dz$, and the latitudinal $O_3$ gradient.

**GEOSCCM** tropical metrics are good, with generally all at 0.5 or higher. The tropical tropopause is slightly colder than observed, but with reasonable water vapour just above the tropical tropopause. Water vapour cannot be analysed for the REF-B2 run in the lower stratosphere, due to problems with the REF-B2 run. In the extra-tropics GEOSCCM is one of the better performing models, however, the amplitude in the NH $H_2O$ seasonal cycle at 200 hPa of the REF-B1 run is too large indicating too large tropospheric influence.

**LMDZrepro** exhibits relatively low temperatures and tropical water vapour. The transport and condensation processes appear to be reasonable, but water vapour is low. Ozone is low at 100hPa as well. LMDZrepro performs reasonably well in the extra-tropics, with reasonable seasonal cycle and meridional gradient in $O_3$ at 200 hPa, while the model achieves lower scores in the metrics testing tropospheric influence such as normalised CO in $d\theta/dz$, the ExTL depth, or the seasonal cycle in $H_2O$ at 200 hPa.

**MRI:** The tropical tropopause temperatures are reasonable, but water vapour is higher than the temperatures would imply due to the presence of ice supersaturation. Ozone

is also higher than observations at 100hPa. MRI performs reasonably well in the extra-tropics, with intermediate to high scores for the dynamical metrics, but relatively low scores for transport and mixing metrics, in particular those testing $H_2O$.

**NiwaSOCOL:** The annual cycle of water vapour in the tropics is shifted slightly. The annual cycle of ozone at 100 hPa is very flat (almost no annual cycle), with too much ozone in NH winter. The tropical tropopause is slightly higher than observations (90 hPa), one level up, though the temperatures are reasonable. NiwaSOCOL is one of the lowest performing models in the extra-tropics. Low scores are obtained in the metrics testing the vertical structure of the tracers, indicating too large tropospheric influence. This is also confirmed by the too large amplitude in the seasonal cycle of $H_2O$ and $O_3$ on the 200 hPa level. NiwaSOCOL it scores reasonably well in the dynamical metrics.

**SOCOL** shows very similar performance to NiwaSOCOL quantitatively and qualitatively in the tropics. It has very large water vapour trends in REF-B2 scenarios, likely due to large trends in tropical tropopause pressure in the 21st century. In the extra-tropics, SOCOL shows similar performance to NiwaSOCOL. The relatively poor performance of NiwaSOCOL and SOCOL in UTLS transport and mixing may stem from the semi-Lagrangian transport scheme, which is known to be overly diffusive. The rather low resolution may enhance the problems due to the transport scheme applied.

**ULAQ** performs acceptably well on tropical metrics except for ozone. Tropical tropopause variability is large. The 100 hPa Ozone annual cycle is large and ozone concentrations are high. Future trends indicate large tropical tropopause temperature and water vapour towards the end of the 21st century. ULAQ is the lowest performing model in the extra-tropics, with a major problem in simulating the latitudinal structure of the zonal-mean zonal wind at 200 hPa. This deficiency is potentially the reason for a very low score of 0.5 for the meridional gradient in $O_3$. Also, the model scores low in reproducing the seasonal cycle in the LMS mass in the SH. The models very low resolution and the use of a geostrophic dynamical core may be the reason for this behaviour.

**UMSLIMCAT:** The tropical tropopause pressure is slightly lower in altitude than other models, but tropical tropopause temperatures are reasonable. The annual cycle of tropical water vapour at 80 hPa and ozone at 100 hPa does not have the same phase over the annual cycle as observed, though tropopause temperatures do have the correct phase. This reduces confidence in TTL transport of water vapour and ozone, though minimum temperature and water vapour correlations are reasonable. In the extra-tropics, UMSLIMCAT is among the better performing models. It achieves the highest score in reproducing the seasonal cycle of $HNO_3$ in the SH and also of $O_3$ in the NH at 200 hPa, pointing towards a good representation of chemistry in the model.

**UMUKCA-METO** has low (120 hPa) tropopause pressure and hence warm CPTs in the tropics. In addition, water vapour has no annual cycle above the cold point, and is low. Tropical water vapour trends in the REF-B2 run seem inconsistent with tropopause temperature trends. Thus there is no correlation between temperature and water vapour. This might be due to errors in processing or the run. In the extra-tropics, UMUKCA-METO is one of the reasonably performing models. Notable deficiencies are a very high LMS mass in both hemispheres, and in the representation of the vertical structure of $O_3$ and $H_2O$, as well as the ExTL depth. The latter is surprising, since the model seems to score highest in the diagnostic testing the CO in $d\theta/dz$, pointing toward a reasonable stratosphere-troposphere exchange.

**UMUKCA-UCAM:** In the tropics, UMUKCA-UCAM is similar in performance to UMUKCA-METO. UMUKCA-UCAM has excessive tropical ozone at 100 hPa. Several diagnostics were not performed due to incorrectly formatted data. In the extra-tropics UMUKCA-UCAM is one of the better performing models, although this might be due to the fact that it is missing some diagnostics. As UMUKCA-METO, it shows too high LMS mass in both hemispheres.

**WACCM** generally has all tropical metrics at or above 0.5. The annual cycle of water vapour is 1-2 months early, but well correlated with temperatures. The 100 hPa ozone annual cycle is a bit flat (less ozone in NH summer). It is one of the best models for CPT. WACCM performs reasonably well in the extra-tropics. The overall model performance is decreased due to a relatively low skill in reproducing the amplitude and phase of the seasonal cycle in $H_2O$ at both 100 and 200 hPa in the NH.

### 7.7.4 Overall Summary

In summary, in the tropical UTLS the models are able to reproduce the climatology of tropopause temperature, pressure, water vapour and ozone, with some common deficiencies. This statement includes both the annual cycle, and interannual anomalies. Interannual anomalies of tropopause pressure are reproduced well. The annual cycle of water vapour in the lower stratosphere is shifted early by a month or more in many models. This indicates likely problems with transport, since tropical cold point tempera-

tures appear to have the correct annual cycle. There is still a large spread in tropical CPTs, but this is smaller than in the CCMVal-1 models. This spread yields a significant spread in stratospheric water vapour, but the transport appears consistent in that models with higher water vapour have higher CPTs, with a few exceptions. Lagrangian cold points can differ substantially from the Eulerian cold point, and biases between models and reanalyses differ. Tropical wave variability is reproduced in the few models examined, but there is also a wide spread in reanalyses. The TIL is found in the tropics in models, but differs from observations. This appears to be partially a function of coarse vertical resolution.

The overall model performance in the extra-tropical UTLS can be summarized as follows: The models score better in metrics testing dynamics (zonal-mean zonal wind and LMS mass) rather than transport and mixing (seasonal cycle of $O_3$ and $H_2O$, meridional gradient in $O_3$, normalised and absolute vertical profiles), and better in metrics focusing on the LS (seasonal cycle of $O_3$ at 100 hPa, meridional gradient in $O_3$) rather than the transition between the troposphere and the stratosphere (seasonal cycle of $O_3$ and $H_2O$ at 200 hPa, and normalised and absolute vertical profiles). This may be simply due to the length-scales of the observed chemical and dynamical structures, which are much smaller in the tropopause region than in the stratosphere. In particular the fine-scale structure of the TIL and ExTL cannot be fully resolved due to the models' limited vertical resolution. The results, however, improve substantially if the models are compared with coarser resolution (or degraded) observations. The multi-model mean generally scores higher than any individual model, except for the seasonal cycle of $H_2O$. In fact, most models seem to score lower in this latter metric, which is likely due to the uncertainty in the observations. The fact that the multi-model mean scores so well on all diagnostics suggests that there are no significant missing processes in the models, although particular models may have significant deficiencies in the representation of the processes.

These results allow a better understanding of model trends, and raise our confidence in the trends. Decreasing observed tropopause pressure trends are highly correlated with observations, especially in the tropics. Trends in extra-tropical tropopause pressure differ between hemispheres due to ozone depletion and recovery. Observed UTLS trends in tropical tropopause temperature are not consistent between reanalyses. Some models show slight decreases in tropopause temperature (observed in some reanalyses), an encouraging sign. These results provide more confidence in future trends. Trends in multiple ensemble members of individual models are similar to each other. CCMVal-2 models predict decreases in tropopause pressure in the tropics and extra-tropics, with different extra-tropical behaviour between hemispheres due to ozone recovery. Tropical tropopause temperatures are expected to increase slightly (1 K per century) with a corresponding consistent increase of 0.5ppmv (about 10-20%) in lower stratospheric water vapour. The magnitude of these trends is consistent across the high performing models.

## References


Anthes, R. A., R. A. Anthes, P. A. Bernhardt, Y. Chen, L. Cucurull, K. F. Dymond, D. Ector, S. B. Healy, S.-P. Ho, D. C. Hunt, Y.-H. Kuo, H. Liu, K. Manning, C. McCormick, T. K. Meehan, W. J. Randel, C. Rocken, W. S. Schreiner, S. V. Sokolovskiy, S. Syndergaard, D. C. Thompson, K. E. Trenberth, T.-K. Wee, N. L. Yen, and Z. Zeng, 2008. The COSMIC/FORMOSAT-3 mission: Early results. *Bull. Amer. Meteor. Soc.*, **89**, 313–333.

Appenzeller, C., J. R. Holton, K. H. Rosenlof, 1996. Seasonal variation of mass transport across the tropopause, *J. Geophys. Res.*, **101**, 15,071–15,078, 10.1029/96JD00821.

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

Bernath, P. F., , C. T. McElroy, M. C. Abrams, C. D. Boone, M. Butler, C. Camy-Peyret, M. Carleer, C. Clerbaux, P.-F. Coheur, R. Colin, P. DeCola, M. DeMazière, J. R. Drummond, D. Dufour, W. F. J. Evans, H. Fast, D. Fussen, K. Gilbert, D. E. Jennings, E. J. Llewellyn, R. P. Lowe, E. Mahieu, J. C. McConnell, M. McHugh, S. D. McLeod, R. Michaud, C. Midwinter, R. Nassar, F. Nichitiu, C. Nowlan, C. P. Rinsland, Y. J. Rochon, N. Rowlands, K. Semeniuk, P. Simon, R. Skelton, J. J. Sloan, M.-A. Soucy, K. Strong, P. Tremblay, D. Turnbull, K. A. Walker, I. Walkty, D. A. Wardle, V. Wehrle, R. Zander, and J. Zou, (2005), Atmospheric Chemistry Experiment (ACE): Mission overview, *Geophys. Res. Lett.*, **32**, doi:10.1029/2005GL022386.

Berthet G., J. G. Esler, P. H. Haynes, 2007. A Lagrangian perspective of the tropopause and the ventilation of the lowermost stratosphere, *J. Geophys. Res.*, **112**, doi:10.1029/2006JD008295.

Birner, T., A. Dornbrack, and U. Schumann, 2002: How sharp is the tropopause at midlatitudes? *Geophys.*

*Res. Lett.*, **29**, doi:10.1029/2002GL015142.

Birner, T., D. Sankey, T. G. Shepherd, 2006. The tropopause inversion layer in models and analyses, *Geophys. Res. Lett.*, **33**, doi:10.1029/2006GL026549.

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

Bönisch, H., Engel, A., Curtius, J., Birner, Th., and Hoor, P., 2009. Quantifying transport into the lowermost stratosphere using simultaneous in-situ measurements of $SF_6$ and $CO_2$, *Atmos. Chem. Phys.*, **9**, 5905-5919.

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

Boone, C. D., R. Nassar, K. A. Walker, Y. Rochon, S. D. McLeod, C. P. Rinsland, and P. F. Bernath, 2005. Retrievals for the Atmospheric Chemistry Experiment Fourier Transform Spectrometer, *Appl. Opt.*, **44**, 7218– 7231.

Bregman, A., J. Lelieveld, M. M. P. van den Broek, P. C. Siegmund, H. Fischer, and O. Bujok, 2000. N2O and O3 relationship in the lowermost stratosphere: A diagnostic for mixing processes as represented by a three-dimensional chemistry-transport model, *J. Geophys. Res.*, **105**, 17,279–17,290.

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

Brunner, D., J. Staehelin, H. L. Rogers, M. O. Köhler, J. A. Pyle, D. Hauglustaine, L. Jourdain, T. K. Berntsen, M. Gauss, I. S. A. Isaksen, E. Meijer, P. van Velthoven, G. Pitari, E. Mancini, G. Grewe, and R. Sausen, 2003. An evaluation of the performance of chemistry transport models by comparison with scientific aircraft observations. Part 1: Concepts and overall model performance., *Atmos. Chem. Phys.*, **3**, 1609–1631.

Brunner, D., J. Staehelin, H. L. Rogers, M. O. Köhler, J. A. Pyle, D. A. Hauglustaine, L. Jourdain, T. K. Berntsen, M. Gauss, I. S. A. Isaksen, E. Meijer, P. van Velthoven, G. Pitari, E. Mancini, V. Grewe, and R. Sausen, 2005. An evaluation of the performance of chemistry transport models - Part 2: Detailed comparison with two selected campaigns, *Atmos. Chem. Phys.*, **5**, 107-129.

Clerbaux, C., M. George, S. Turquety, K. A. Walker, B. Barret, P. Bernath, C. Boone, T. Borsdorff, J. P. Cammas, V. Catoire, M. Coffey, P.-F. Coheur, M. Deeter, M. De Mazière, J. Drummond, P. Duchatelet, E. Dupuy, R. de Zafra, F. Eddounia, D. P. Edwards, L. Emmons, B. Funke, J. Gille, D. W. T. Griffith, J. Hannigan, F. Hase, M. Höpfner, N. Jones, A. Kagawa, Y. Kasai, I. Kramer, E. Le Flochmoën, N. J. Livesey, M. López-Puertas, M. Luo, E. Mahieu, D. Murtagh, P. Nédélec, A. Pazmino, H. Pumphrey, P. Ricaud, C. P. Rinsland, C. Robert, M. Schneider, C. Senten, G. Stiller, A. Strandberg, K. Strong, R. Sussmann, V. Thouret, J. Urban, and A. Wiacek, 2008. CO measurements from the ACE-FTS satellite instrument: Data analysis and validation using ground-based, airborne and spaceborne observations, *Atmos. Chem. Phys.*, **8**, 2569–2594.

Considine, D. B., M. Natarajan, T. D. Fairlie, G. S. Lingenfelser, R. B. Pierce, L. Froidevaux, and A. Lambert, 2008. Noncoincident validation of Aura MLS observations using the Langley Research Center Lagrangian chemistry and transport model, *J. Geophys. Res.*, **113**, doi:10.1029/2007JD008770.

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

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

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

Dupuy, E., K. A. Walker, J. Kar, C. D. Boone, C. T. McElroy, P. F. Bernath, J. R. Drummond, R. Skelton, S. D. McLeod, R. C. Hughes, C. R. Nowlan, D. G. Dufour, J. Zou, F. Nichitiu, K. Strong, P. Baron, R. M. Bevilacqua, T. Blumenstock, G. E. Bodeker, T. Borsdorff, A. E. Bourassa, H. Bovensmann, I. S. Boyd, A. Bracher, C. Brogniez, J. P. Burrows, V. Catoire, S. Ceccherini, S. Chabrillat, T. Christensen, M. T. Coffey, U. Cortesi, J. Davies, C. De Clercq, D. A. Degenstein, M. De Mazière, P. Demoulin, J. Dodion,

B. Firanski, H. Fischer, G. Forbes, L. Froidevaux, D. Fussen, P. Gerard, S. Godin-Beekmann, F. Goutail, J. Granville, D. Griffith, C. S. Haley, J. W. Hannigan, M. Höpfner, J. J. Jin, A. Jones, N. B. Jones, K. Jucks, A. Kagawa, Y. Kasai, T. E. Kerzenmacher, A. Kleinböhl, A. R. Klekociuk, I. Kramer, H. Küllmann, J. Kuttippurath, E. Kyrölä, J.-C. Lambert, N. J. Livesey, E. J. Llewellyn, N. D. Lloyd, E. Mahieu, G. L. Manney, B. T. Marshall, J. C. McConnell, M. P. McCormick, I. S. McDermid, M. McHugh, C. A. McLinden, J. Mellqvist, K. Mizutani, Y. Murayama, D. P. Murtagh, H. Oelhaf, A. Parrish, S. V. Petelina, C. Piccolo, J.-P. Pommereau, C. E. Randall, C. Robert, C. Roth, M. Schneider, C. Senten, T. Steck, A. Strandberg, K. B. Strawbridge, R. Sussmann, D. P. J. Swart, D. W. Tarasick, J. R. Taylor, C. Tétard, L. W. Thomason, A. M. Thompson, M. B. Tully, J. Urban, F. Vanhellemont, C. Vigouroux, T. von Clarmann, P. von der Gathen, C. von Savigny, J. W. Waters, J. C. Witte, M. Wolff, and J. M. Zawodny, 2009. Validation of ozone measurements from the Atmospheric Chemistry Experiment (ACE), *Atmos. Chem. Phys.*, **9**, 287-343.

Engel, A., H. Bönisch, D. Brunner, H. Fischer, H. Franke, G. Günther, C. Gurk, M. Hegglin, P. Hoor, R. Königstedt, M. Krebsbach, R. Maser, U. Parchatka, T. Peter, D. Schell, C. Schiller, U. Schmidt, N. Spelten, T. Szabo, U. Weers, H. Wernli, T. Wetter, and V. Wirth, 2006. Highly resolved observations of trace gases in the lowermost stratosphere and upper troposphere from the SPURT project: An overview, *Atmos. Chem. Phys.*, **6**, 283– 301.

Eyring, V., N. Butchart, D. W. Waugh, H. Akiyoshi, J. Austin, S. Bekki, G. E. Bodeker, B. A. Boville, C. Brühl, M. P. Chipperfield, E. Cordero, M. Dameris, M. Deushi, V. E. Fioletov, S. M. Frith, R. R. Garcia, A. Gettelman, M. A. Giorgetta, V. Grewe, L. Jourdain, D. E. Kinnison, E. Mancini, E. Manzini, M. Marchand, D. R. Marsh, T. Nagashima, P. A. Newman, J. E. Nielsen, S. Pawson, G. Pitari, D. A. Plummer, E. Rozanov, M. Schraner, T. G. Shepherd, K. Shibata, R. S. Stolarski, H. Struthers, W. Tian, and M. Yoshiki, 2006. Assessment of temperature, trace species, and ozone in chemistry-climate model simulations of the recent past, *J. Geophys. Res.*, **111**, doi:10.1029/2006JD007327.

Fischer, H., M. Birk, C. Blom, B. Carli, M. Carlotti, T. von Clarmann, L. Delbouille, A. Dudhia, D. Ehhalt, M. Endemann, J. M. Flaud, R. Gessner, A. Kleinert, R. Koopman, J. Langen, M. López-Puertas, P. Mosner, H. Nett, H. Oelhaf, G. Perron, J. Remedios, M. Ridolfi, G. Stiller, and R. Zander, 2008. MIPAS: an instrument for atmospheric and climate research, *Atmos. Chem. Phys.*, **8**, 2151-2188.

Flocke, F., R. L. Herman, R. J. Salawitch, E. Atlas, C. R. Webster, S. M. Schauffler, R. A. Lueb, R. D. May, E. J. Moyer, K. H. Rosenlof, D. C. Scott, D. R. Blake, and T. P. Bui, 1999. An examination of chemistry and transport processes in the tropical lower stratosphere using observations of long-lived and short-lived compounds obtained during STRAT and POLARIS, *J. Geophys. Res.*, **104**, 26,625– 26,642.

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

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

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

Fujiwara, M., and M. Takahashi, 2001. Role of the equatorial Kelvin wave in stratosphere-troposphere exchange in a general circulation model, *J. Geophys. Res.*, **106**, 22,763– 22,780.

Fujiwara, M., S. Iwasaki, A. Shimizu, Y. Inai, M. Shiotani, F. Hasebe, I. Matsui, N. Sugimoto, H. Okamoto, N. Nishi, A. Hamada, T. Sakazaki, and K. Yoneyama, 2009. Cirrus observations in the tropical tropopause layer over the western Pacific, *J. Geophys. Res.*, **114**, doi:10.1029/2008JD011040.

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

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

Gettelman, A., P. M. F. Forster, M. Fujiwara, Q. Fu, H. Vomel, L. K. Gohar, C. Johanson and M. Ammeraman, 2004. The Radiation Balance of the

Tropical Tropopause Layer, *J. Geophys. Res.*, **109**, doi:10.1029/2003JD004190.

Gettelman, A., T. Birner, V. Eyring, H. Akiyoshi, S. Bekki, C. Brühl, M. Dameris, D. E. Kinnison, F. Lefevre, F. Lott, E. Mancini, G. Pitari, D. A. Plummer, E. Rozanov, K. Shibata, A. Stenke, H. Struthers, and W. Tian, 2009. The tropical tropopause 1960-2100. *Atmos. Chem. Phys.*, **9**, 1621–1637.

Gettelman, A., M. I. Hegglin, S.-W. Son, J. Kim, M. Fujiwara, T. Birner, S. Kremser, M. Rex, J. A. Anel, J. Austin, J.-F. Lamarque, H. Akiyoshi, D. Plummer, T. G. Shepherd, J. Scinocca, M. Michou, H. Teyssèdre, M. Dameris, H. Garny, C. Brühl, P. Jöckel, S. Pawson, S. Bekki, K. Shibata, D. Smale, E. Rozanov, E. Mancini, G. Pitari, O. Morgenstern, M. Chipperfield, S. Dohmse, W. Tian, N. Butchart, S. C. Hardiman, P. Braesicke, J. A. Pyle, D. E. Kinnison, 2010. Multimodel assessment of the upper troposphere and lower stratosphere part II: Tropics and global trends, *J. Geophys. Res.*, doi:10.1029/2009JD013638, in press.

Gilman, D. L., F. J. Fuglister, and J. M. Mitchell Jr., 1963. On the power spectrum of "red noise," *J. Atmos. Sci.*, **20**, 182–184.

Grewe, V., and R. Sausen, 2009. Comment on "quantitative performance metrics for stratospheric-resolving chemistry-climate models" by Waugh and Eyring, *Atmos. Chem. Phys.*, **9**, 9101–9110.

Grise, K. M., D. W. J. Thompson, and T. Birner, 2010. A global survey of static stability, *J. Clim.*, doi: 10.1175/2009JCLI3369.1, in press.

Hassler, B., G. E. Bodeker, and M. Dameris, 2008. 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.

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

Hegglin, M. I., D. Brunner, T. Peter, J. Staehelin, V. Wirth, P. Hoor, and H. Fischer, 2005. Determination of eddy-diffusivity in the lowermost stratosphere, *Geophys. Res. Lett.*, **32**, doi:10.1029/2005GL022495.

Hegglin, M. I., D. Brunner, T. Peter, P. Hoor, H. Fischer, J. Staehelin, M. Krebsbach, C. Schiller, U. Parchatka, and U. Weers, 2006. Measurements of NO, $NO_y$, $N_2O$, and $O_3$ during SPURT: implications for transport and chemistry in the lowermost stratosphere, *Atmos. Chem. Phys.*, **6**, 1331–1350.

Hegglin, M. I., and Shepherd, T. G., 2007. $O_3$-$N_2O$ correlations from the Atmospheric Chemistry Experiment: Revisiting a diagnostic of transport and chemistry in the stratosphere, *J. Geophys. Res.*, **112**, doi:10.1029/2006JD008281.

Hegglin, M. I., C. D. Boone, G. L. Manney, T. G. Shepherd, K. A. Walker, P. F. Bernath, W. H. Daffer, P. Hoor, and C. Schiller, 2008. Validation of ACE-FTS satellite data in the upper troposphere/lower stratosphere (UTLS) using non-coincident measurements, *Atmos. Chem. Phys.*, **8**, 1483–1499.

Hegglin, M. I., C. D. Boone, G. L. Manney, and K. A. Walker, 2009. A global view of the extratropical tropopause transition layer from Atmospheric Chemistry Experiment Fourier Transform Spectrometer $O_3$, $H_2O$, and CO, *J. Geophys. Res.*, **114**, doi:10.1029/2008JD009984.

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

Hegglin, M. I., A. Gettelman, P. Hoor, R. Krichevsky, G. L. Manney, L. L. Pan, S-W. Son, G. Stiller, S. Tilmes, K. A. Walker, V. Eyring, T. G. Shepherd, D. Waugh, H. Akiyoshi, J. Austin, A. Baumgaertner, S. Bekki, P. Braesicke, C. Brühl, N. Butchart, M. Chipperfield, M. Dameris, S. Dhomse, S. Frith, H. Garny, S. C. Hardiman, P. Jöckel, D. E. Kinnison, J. F. Lamarque, E. Mancini, M. Michou, O. Morgenstern, T. Nakamura, D. Olivié, S. Pawson, G. Pitari, D. A. Plummer, E. Rozanov, J. F. Scinocca, K. Shibata, D. Smale, H. Teyssèdre, W. Tian, Y. Yamashita, 2010. Multi-Model Assessment of the Upper Troposphere and Lower Stratosphere part I: Extra-tropics, *J. Geophys. Res.*, doi:10.1029/2010JD013884, submitted.

Highwood, E. J., and B. J. Hoskins, 1998. The tropical tropopause, *Quart. J. Roy. Meteorol. Soc.*, **124**, 1579–1604.

Hintsa, E. J., E. M. Weinstock, J. G. Anderson, R. D. May, and D. F. Hurst, 1999. On the accuracy of in situ water vapor measurements in the troposphere and lower stratosphere with the Harvard Lyman-a hygrometer, *J. Geophys. Res.*, **104**, 8183– 8189.

Holton, J., P. Haynes, M. Mclntyre, A. Douglass, R. Rood, and L. Pfister, 1995. Stratosphere-troposphere exchange. *Rev. Geophys.,* **33**, 403–439.

Hoor, P., Fischer, H., Lange, L., Lelieveld, J., Brunner, D., 2002. Seasonal variations of a mixing layer in the lowermost stratosphere as identified by the CO-$O_3$ correlation from in situ measurements, *J. Geophys. Res.*, **107**, doi:10.1029/2000JD000289.

Hoor, P., Brunner, D., Hegglin, M., Peter, Th., Gurk, C., Wernli, H., Fischer, H., 2004. Seasonality and extent of extra-tropical TST derived from in-situ CO measurements during SPURT, *Atmos. Chem. Phys.*, **4**, 1427–1442.

Hoor, P., H. Fischer, and J. Lelieveld, 2005. Tropical and extra-tropical tropospheric air in the lowermost stratosphere over Europe: A CO-based budget, *Geophys. Res. Lett.*, **32**, doi:10.1029/2004GL022018.

Houghton, J. T., Y. Ding, D.J. Griggs, M. Noguer, P.J. van der Linden, X. Dai, K. Maskell, and C.A. Johnson, editors. Climate Change 2001: The Scientific Basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, UK, 2001.

Kalnay, E., M. Kanamitsu, R. Kistler, W. Collins, D. Deaven, L. Gandin, M. Iredell, S. Saha, G. White, J. Woollen Y. Zhu, M. Chelliah, W. Ebisuzaki, W. Higgins, J. Janowiak, K. C. Mo, C. Ropelewski, J. Wang, A. Leetmaa, R. Reynolds, R. Jenne, and D. Joseph, 1996. The NCEP/NCAR 40-year reanalysis project, *Bull. Am. Meteorol. Soc.*, **77**, 437–471.

Kanamitsu, M., W. Ebisuzaki, J. Wollen, S.-K. Yang, J. J. Hnilo, M. Fiorino, and G. L. Potter, 2002. NCEP-DEO AMIP II reanalysis (R-2), *Bull. Am. Meteorol. Soc.*, **83**, 1631–1643.

Kley, D., J. M. Russell III, and C. Phillips (Eds.), 2000. SPARC Assessment of Upper Tropospheric and Stratospheric Water Vapour, SPARC Report No 2, WCRP Report 113, WMO/TD-N° 1043.

Kremser, S., I. Wohltmann, M. Rex, U. Langematz, M. Dameris, M. Kunze, 2009. Water vapor transport in the tropical tropopause region in couple chemistry-climate models and ERA-40 reanalysis data, *Atmos. Chem. Phys.*, **9**, 2679-2694.

Levine, J. G., P. Braesicke, N. R. P. Harris, N. H. Savage, and J. A. Pyle, 2007. Pathways and timescales for troposphere-tostratosphere transport via the tropical tropopause layer and their relevance for very short lived substances, *J. Geophys. Res.*, **112**, doi:10.1029/2005JD006940.

Levine, J. G., P. Braesicke, N. R. P. Harris, and J. A. Pyle, 2008. Seasonal and inter-annual variations in troposphere-to-stratosphere transport from the tropical tropopause layer, *Atmos. Chem. Phys.*, **8**, 3689–3703.

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

Livesey, N. J. et al., 2007. MLS Version 2.2 Level 2 data quality and description document, Tech. Rep. JPL D-33509, Jet Propulsion Laboratory, available at http://mls.jpl.nasa.gov

Livesey, N. J., M. J. Filipiak, L. Froidevaux, W. G. Read, A. Lambert, M. L. Santee, J. H. Jiang, H. C. Pumphrey, J. W. Waters, R. E. Cofield, D. T. Cuddy, W. H. Daffer, B. J. Drouin, R. A. Fuller, R. F. Jarnot, Y. B. Jiang, B. W. Knosp, Q. B. Li, V. S. Perun, M. J. Schwartz, W. V. Snyder, P. C. Stek, R. P. Thurstans, P. A. Wagner, M. Avery, E. V. Browell, J.-P. Cammas, L. E. Christensen, G. S. Diskin, R.-S. Gao, H.-J. Jost, M. Loewenstein, J. D. Lopez, P. Nedelec, G. B. Osterman, G. W. Sachse, an d C. R. Webster, 2008. Validation of Aura Microwave Limb Sounder $O_3$ and CO observations in the upper troposphere and lower stratosphere, *J. Geophys. Res.*, **113**, doi:10.1029/2007JD008805.

Logan, J., 1999. An analysis of ozonesonde data for the lower stratosphere: Recommendations for testing models, *J. Geophys. Res.*, **104**, 16,151–16,170.

Matsuno, T., 1966. Quasi-geostrophic motions in the equatorial area, *J. Meteor. Soc. Japan*, **44**, 25–43.

Milz, M., T. von Clarmann, H. Fischer, N. Glatthor, U. Grabowski, M. Höpfner, S. Kellmann, M. Kiefer, A. Linden, G. Mengistu Tsidu, T. Steck, G. P. Stiller, B. Funke, M. López-Puertas, and M. E. Koukouli, 2005. Water vapor distributions measured with the Michelson Interferometer for Passive Atmospheric Sounding on board Envisat (MIPAS/Envisat), *J. Geophys. Res.*, **110**, doi:10.1029/2005JD005973.

Milz, M., T. v. Clarmann, P. Bernath, C. Boone, S. A. Buehler, S. Chauhan, B. Deuber, D. G. Feist, B. Funke, N. Glatthor, U. Grabowski, A. Griesfeller, A. Haefele, M. Höpfner, N. Kämpfer, S. Kellmann, A.

Linden, S. Müller, H. Nakajima, H. Oelhaf, E. Remsberg, S. Rohs, J. M. Russell III, C. Schiller, G. P. Stiller, T. Sugita, T. Tanaka, H. Vömel, K. Walker, G. Wetzel, T. Yokota, V. Yushkov, and G. Zhang, 2009. Validation of water vapour profiles (version 13) retrieved by the IMK/IAA scientific retrieval processor based on full resolution spectra measured by MIPAS on board Envisat, *Atmos. Meas. Tech.*, **2**, 379–399.

Milosevich, L. M., A. Paukkunen, H. Vömel, S. J. Oltmans, 2004. Development and Validation of a Time-Lag Correction for Vaisala Radiosonde Humidity Measurements, *J. Atmos. Ocean Tech.*, **21**, 1305–1327.

Mote, P. W., K. H. Rosenlof, M. E. Mclntyre, E. S. Carr, J. C. Gille, J. R. Holton, J. S. Kinnersley, H. C. Pumphrey, J. M. Russell, and J. W. Waters, 1996. An atmospheric tape recorder: The imprint of tropical tropopause temperatures on stratospheric water vapor. *J. Geophys. Res.*, **101**, 3989–4006.

Newman, P. A., D. W. Fahey, W. H. Brune, and M. J. Kurylo, 1999. Preface to special section: Photochemistry of Ozone Loss in the Arctic Region in Summer (POLARIS), *J. Geophys. Res.*, **104**, 26,481–26,495.

Onogi, K., J. Tsutsui, H. Koide, M. Sakamoto, S. Kobayashi, H. Hatsushika, T. Matsumoto, N. Yamazaki, H. Kamahori, K. Takahashi, S. Kadokura, K. Wada, K. Kato, R. Oyama, T. Ose, N. Mannoji and R. Taira, 2007. The JRA-25 Reanalysis. *J. Meteor. Soc. Japan*, **85**, 369–432.

Pan, L.L., W. J. Randel, B. L. Gary, M. J. Mahoney, and E. J. Hintsa, 2004. Definitions and sharpness of the extra-tropical tropopause: A trace gas perspective, *J. Geophys. Res.*, **109**, doi:10.1029/2004JD004982.

Pan, L. L., J. C. Wei, D. E. Kinnison, R. R. Garcia, D. J. Wuebbles, and G. P. Brasseur, 2007. A set of diagnostics for evaluating chemistry-climate models in the extra-tropical tropopause region, *J. Geophys. Res.*, **112**, doi:10.1029/2006JD007792.

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

Proffitt, M. H., and R. J. McLaughlin, 1983. Fast-response dual-beam UV-absorption ozone photometer suitable for use on stratospheric balloons, *Rev. Sci. Instrum.*, **54**, 1719–1728.

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

Randel, W. J., F. Wu, and P. Forster, 2007. The Extratropical Tropopause Inversion Layer: Global Observations with GPS Data, and a Radiative Forcing Mechanism, *J. Atmos. Sci.*, **64**, 4489–4496.

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

Ray, E. A., Moore, F. L., Elkins, J. W., Dutton, G. S., Fahey, D. W., Vömel, H., Oltmans, S. J., and Rosenlof, K. H., 1999. Transport into the Northern Hemisphere lowermost stratosphere revealed by in situ tracer measurements, *J. Geophys. Res.*, **104**, 26,565–26,580.

Rood R. B., A. Douglass, M. Cerniglia, L. Sparling, and J. Nielsen, 2000. Seasonal variability of middle-latitude ozone in the lowermost stratosphere derived from probability distribution functions, *J. Geophys. Res.*, **105**, 17,793–17,805.

Rosenlof, K., A. F. Tuck, K. K. Kelly, J. M. Russell, and M. P. McCormick, 1997. Hemispheric asymmetries in water vapor and inferences about transport in the lower stratosphere, *J. Geophys. Res.*, **102**, 13,213–13,234.

Russell, J. M., L. L. Gordley, J. H. Park, S. R. Drayson, A. F. Tuck, J. E. Harries, R. J. Cicerone, P. J. Crutzen, and J. E. Frederick, 1993. The Halogen Occultation Experiment, *J. Geophys. Res.*, **98**, 10,777–10,797.

Schoeberl, M. R, A. R. Douglass, Z. Zhu and S. Pawson, 2003. A comparison of the lower stratospheric age spectra derived from a general circulation model and two data assimilation systems, *J. Geophys. Res.*, **108**, doi:10.1029/2002JD002652.

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

Shepherd, T. G., 2002. Issues in Stratosphere-troposphere Coupling, *J. Meteor. Soc. Japan,* **80**, 769.

Shepherd, T. G., 2007. Transport in the Middle Atmosphere, *J. Meteor. Soc. Japan*, **85B**, 165–191.

Son, S. W., L. M. Polvani, D. W. Waugh, T. Birner, , H. Akiyoshi, R. R. Garcia, A. Gettelman, D. A. Plummer, and E. Rozanov, 2009a. The impact of stratospheric ozone recovery on tropopause height trends. *J. Clim.*, **22**, 429–445.

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

Steck, T., T. von Clarmann, H. Fischer, B. Funke, N. Glatthor, U. Grabowski, M. Höpfner, S. Kellmann, M. Kiefer, A. Linden, M. Milz, G. P. Stiller, D. Y. Wang, M. Allaart, Th. Blumenstock, P. von der Gathen, G. Hansen, F. Hase, G. Hochschild, G. Kopp, E. Kyrö, H. Oelhaf, U. Raffalski, A. Redondas Marrero, E. Remsberg, J. Russell III, K. Stebel, W. Steinbrecht, G. Wetzel, M. Yela, and G. Zhang, 2007. Bias determination and precision validation of ozone profiles from MIPAS-Envisat retrieved with the IMK-IAA processor, *Atmos. Chem. Phys.*, **7**, 3639–3662.

Suzuki, J., and M. Shiotani, 2008. Space-time variability of equatorial Kelvin waves and intraseasonal oscillations around the tropical tropopause, *J. Geophys. Res.*, **113**, doi:10.1029/2007JD009456.

Taylor, K. E., 2001. Summarizing multiple aspects of model performance in a single diagram, *J. Geophys. Res.*, **106**, 7183–7192.

Tilmes, S., L. Pan, P. Hoor, G. W. Sachse, M. Loewenstein, J. Lopez, C. Webster, L. E. Cristensen, M. Proffitt, R.-S. Gao, G. S. Diskin, M. A. Avery, J. R. Podolske, R. L. Herman, N. Spelten, A. Weinheimer, T. Campus, E. J. Hintsa, E. M. Weinstock, J. Pittman, M. A. Zondl, M. E. Paige, E. Atlas, 2010. An aircraft based upper troposphere lower stratosphere $O_3$, CO and $H_2O$ climatology for the Northern hemisphere, *J. Geophys. Res.*, doi:10.1029/2009JD012731.

Tsuda, T., Y. Murayama, H. Wiryosumarto, S. Harijono, and S. Kato, 1994. Radiosonde observations of equatorial atmosphere dynamics over Indonesia: 1. Equatorial waves and diurnal tides, *J. Geophys. Res.*, **99**, 10,491–10,505.

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. M., P. W. Kållberg, A. J. Simmons, U. Andrae, V. da Costa Bechtold, M. Fiorino, K. K. Gibson, J. Haseler, A. Hernandez, G. A. Kelly, X. Li, K. Onogi, S. Saarinen, N. Sokka, R. P. Allan, E. Andersson, K. Arpe, M. A. Balmaseda, A. C. M. Beljaars, L. van de Berg, J. Bidlot, N. Bormann, S. Caires, F. Chevallier, A. Dethof, M. Dragosavac, M. Fisher, M. Fuentes, S. Hagemann, E. Hólm, B.J. Hoskins, L. Isaksen, P. A. E. M. Janssen, R. Jenne, A. P. McNally, J.-F. Mahfouf, J.-J. Morcrette, N. A. Rayner, R.W. Saunders, P. Simon, A. Sterl, K. E. Trenberth, A. Untch, D. Vasiljevic, P. Viterbo, and J. Woollen, 2005. The ERA-40 re-analysis, *Quart. J. Roy. Meteorol. Soc.*, **131**, 2961–3012.

von Clarmann, T., N. Glatthor, U. Grabowski, M. Höpfner, S. Kellmann, M. Kiefer, A. Linden, G. M. Tsidu, M. Milz, T. Steck, G. P. Stiller, D. Y. Wang, H. Fischer, B. Funke, S. Gil-López, M. López-Puertas, 2003. Retrieval of temperature and tangent altitude pointing from limb emission spectra recorded from space by the Michelson Interferometer for Passive Atmospheric Sounding (MIPAS), *J. Geophys. Res.*, **108**, doi:10.1029/2003JD003602.

von Clarmann, T., M. Höpfner, S. Kellmann, A. Linden, S. Chauhan, B. Funke, U. Grabowski, N. Glatthor, M. Kiefer, T. Schieferdecker, G. P. Stiller, and S. Versick, 2009. Retrieval of temperature, $H_2O$, $O_3$, $HNO_3$, $CH_4$, $N_2O$, $ClONO_2$ and ClO from MIPAS reduced resolution nominal mode limb emission measurements, *Atmos. Meas. Tech.*, **2**, 159–175.

Wang, D. Y., M. Höpfner, G. Mengistu Tsidu, G. P. Stiller, T. von Clarmann, H. Fischer, T. Blumenstock, N. Glatthor, U. Grabowski, F. Hase, S. Kellmann, A. Linden, M. Milz, H. Oelhaf, M. Schneider, T. Steck, G. Wetzel, M. López-Puertas, B. Funke, M. E. Koukouli, H. Nakajima, T. Sugita, H. Irie, J. Urban, D. Murtagh, M. L. Santee, G. Toon, M. R. Gunson, F. W. Irion, C. D. Boone, K. Walker, and P. F. Bernath, 2007. Validation of nitric acid retrieved by the IMK-IAA processor from MIPAS/ENVISAT measurements, *Atmos. Chem. Phys.*, **7**, 721–738.

Waters, J. W., Froidevaux, L., Harwood, R.S., Jarnot, R.F., Pickett, H.M., Read, W.G., Siegel, P.H., Cofield, R.E., Filipiak, M.J., Flower, D.A., Holden, J.R. Lau, G.K., Livesey, N.J., Manney, G.L., Pumphrey, H.C., Santee, M.L., Wu, D.L., Cuddy, D.T., Lay, R.R., Loo, M.S., Perun, V.S., Schwartz, M.J.,

Stek, P.C., Thurstans, R.P., Boyles, M.A., Chandra, K.M., Chavez, M.C., Gun-Shing Chen, Chudasama, B.V., Dodge, R., Fuller, R.A., Girard, M.A., Jiang, J.H., Yibo Jiang, Knosp, B.W., LaBelle, R.C., Lam, J.C., Lee, K.A., Miller, D., Oswald, J.E., Patel, N.C., Pukala, D.M., Quintero, O., Scaff, D.M., Van Snyder, W., Tope, M.C., Wagner, P.A., and Walch, M.J., 2006. The Earth Observing System Microwave Limb Sounder (EOS MLS) on the Aura satellite, *IEEE Trans. Geosci. Remote Sens.*, **44**, 1075–1092.

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

Wheeler, M., and G. N. Kiladis, 1999. Convectively coupled equatorial waves: Analysis of clouds and temperature in the wavenumber-frequency domain. *J. Atmos. Sci.*, **56**, 374–399.

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

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