Chapter 5

Transport

Lead Authors: Jessica Neu & Susan Strahan

Co-Authors: Peter Braesicke
Anne Douglass
Petra Huck
Luke Oman
Diane Pendlebury
Susann Tegtmeier

5.1 Introduction

The distribution of long-lived trace gases in the stratospheric overworld is controlled mainly by the balance between the diabatic circulation, which acts to create equator-to-pole gradients in tracer isopleths, and quasi-horizontal mixing, which acts to flatten tracer isopleths in mixing regions while sharpening gradients at the locations of mixing barriers. Three important barriers to transport are the subtropical barrier (*i.e.*, the tropical "pipe"), the edge of the polar vortex, and the extra-tropical tropopause. A schematic of the stratospheric circulation and transport barriers is shown in **Figure 5.1**. Both the strength of the diabatic circulation and that of the transport barriers are linked to wave activity in the stratosphere and thus vary with height and season.

In this chapter, model representation of stratospheric transport processes is evaluated using process-oriented diagnostics derived from observations. In many cases it is impossible to create a diagnostic that uniquely assesses a single process because most have contributions from multiple transport mechanisms. For example, the age of air in the tropics depends on integrated vertical advection, vertical mixing, and mixing between the tropics and extra-tropics across the subtropical transport barrier. However, combining the diagnostics and, where possible, using the information from diagnostics that do uniquely assess a

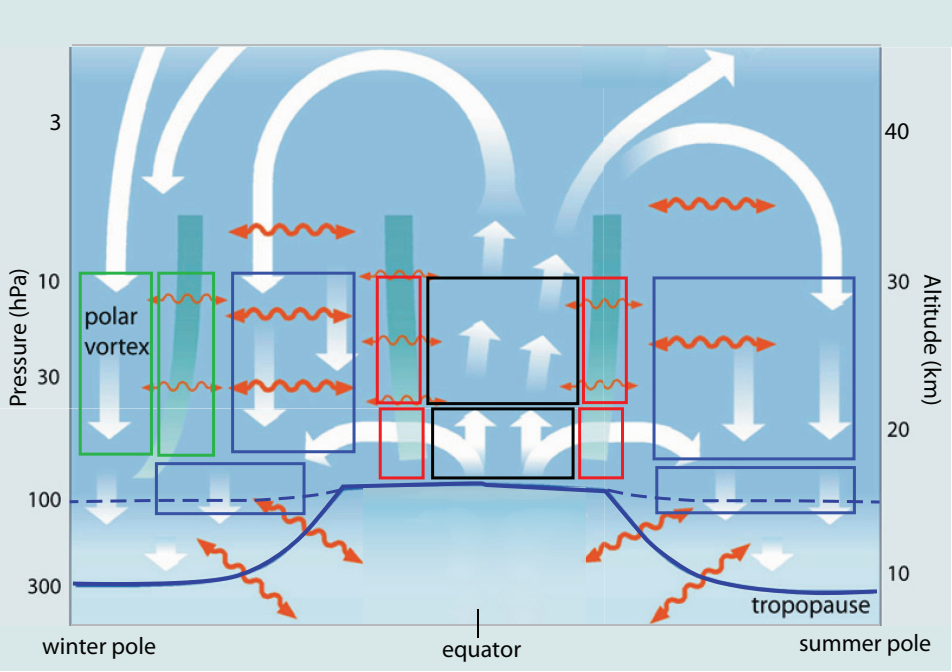

Figure 5.1: Schematic of the stratospheric circulation (courtesy of Harald Boenisch). The blue dashed line shows the boundary between the overworld and the lowermost stratosphere (380 K = ~100 hPa, 16 km). White arrows indicate the direction of the residual circulation in the winter and summer hemispheres. Transport barriers are shown as dark blue shaded regions. Red two-headed arrows indicate mixing, with smaller arrows showing weaker mixing across transport barriers. The boxes indicate regions where transport processes are evaluated in this chapter, with the colour of the box corresponding to the Region/Process column of Table 5.1 and the chapter sections (Black = Tropical ascent, Red = Tropical-extra-tropical mixing, Blue = Integrated processes affecting mid-latitude composition, Green = Polar processes) The lowermost stratosphere is not evaluated in this chapter. A complete list of transport diagnostics can be found in Table 5.1.

single process allows model behaviour to be broken down and assessed in a physically-meaningful way. The suite of diagnostics chosen for this chapter (**Table 5.1**) is intended to cover the major processes controlling stratospheric trace gas distributions in the overworld (above 380 K, ~100 hPa), from entry in the tropical lower stratospheric to exit through the extra-tropical 100 hPa surface. The diagnostics emphasize transport processes below 10 hPa because the goal of this report is to understand 21st century O_3 in the WMO reference simulations. The present day (~1990-2006) behaviour of CCMs is evaluated and compared to observations, using the REF-B1 simulations for all models except UMETRAC. For UMETRAC, REF-B0, the present day time slice simulation, was used because REF-B1 was unavailable. Diagnostics are calculated from 10-15 years of model output whenever possible. Future changes in a few key diagnostics are presented by comparing changes between the recent past (1990-2006) and the future (2080-2099) from the REF-B2 simulations.

More than a dozen diagnostics are applied here but many have overlap in the processes they evaluate. The results presented in this chapter are grouped by process and region in the same order as listed in Table 5.1. The processes evaluated include tropical ascent, mixing between the tropics and mid-latitudes (*i.e.*, the "leakiness" of the subtropical barriers), and descent and isolation in the polar region. As a connection to the upper troposphere/lower stratosphere (UTLS) region evaluated in Chapter 7, the influence of seasonal variations in transport on air leaving the stratosphere is also examined. Unless otherwise noted, model performance on the diagnostics presented here has been quantified using the metric described in Waugh and Eyring (2008), hereafter referred to as WE08, using 3σ of the observational uncertainty in the denominator. Because there are significant differences in model performance between the lower ($\geq$ 50 hPa) and middle stratosphere ($\leq$ 50 hPa) (LS and MS, respectively), many of the diagnostics are applied separately to the two regions.

5.2 Transport Diagnostics for the Tropics

Most air enters the stratosphere in the tropics, so transport in this region is critical to determining stratospheric composition. Only by simulating the correct balance be-

Table 5.1: Stratospheric Transport Diagnostics for CCMs. Gray highlights indicate diagnostics that are used as quantitative metrics for the overall model performance.

Diagnostic	Variables	Observations	References
Tropical Ascent			
H_2O tape recorder phase	H_2O and CH_4	UARS HALOE and Aura MLS	*Mote et al. (1996; 1998); Hall et al. (1999); Schoeberl et al. (2008)*
Tropical-Mid-latitude Mean Age Gradient	CO_2 and SF_6	Various balloon missions	*Andrews et al. (2001); Engel et al. (2009)*
Tropical-Extratropical Mixing			
H_2O tape recorder amplitude	H_2O and CH_4	UARS HALOE and Aura MLS	*Mote et al. (1996; 1998); Hall et al. (1999); Schoeberl et al. (2008)*
Tropical Mean Age profile	CO_2 and SF_6	OMS balloon profiles	*Andrews et al. (2001)*
Tropical CH_4 vertical gradient	CH_4	UARS HALOE	*Eyring et al. (2006)*
Tropical-Mid-latitude N_2O PDFs	N_2O, potential temperature	ENVISAT-MIPAS and Aura MLS	*Douglass et al. (1999); Sparling (2000)*
Integrated Processes Affecting Extratropical Composition			
NH Mid-latitude Mean Age profile	CO_2 and SF_6	Various balloon missions	*Engel et al. (2009)*
Fractional Release of Cly	CFC-11, CFC-12, CO_2, and SF_6	NASA ER-2 aircraft missions	*Schauffler et al. (2003); Douglass et al. (2008)*
NH Mid-latitude Cl_y time series	Cl_y	UARS HALOE and Aura MLS	*Lary et al. (2007); Waugh and Eyring (2008)*
N_2O annual cycle in the LS	N_2O	Aura MLS	
Mean Age at 60°N/S	CO_2	NASA ER-2 aircraft missions	*Andrews et al. (2001); Boering et al. (1996)*
Polar Processes			
Antarctic spring CH_4 PDFs	CH_4, potential temperature	UARS HALOE	*Strahan and Polansky (2006)*
Antarctic September N_2O profile	N_2O	Aura MLS	
Antarctic spring Cl_y time series	Cl_y	UARS HALOE and Aura MLS	*Douglass et al. (1995); Santee et al. (1996); Lary et al. (2007)*

tween ascent and mixing across the subtropical barriers can a model get the correct balance of pathways necessary for accurate simulation of photochemically active species.

5.2.1 Ascent

5.2.1.1 Tape Recorder Phase Speed

Air entering the stratosphere through the tropics slowly ascends with limited horizontal mixing. The vertical propagation of annual variations in tropical tropopause water vapour, known as the "tape recorder" signal (*e.g.*, Mote *et al.*, 1998; Hall *et al.*, 1999, Waugh and Hall, 2002, Schoeberl *et al.*, 2008), provides striking visual evidence of this isolated ascent. **Figure 5.2** shows the deviation of the water vapour mixing ratio from the monthly mean profile averaged over 10°S to 10°N for combined HALOE and MLS observations, and for the available models (data courtesy of Mark Schoeberl). UMETRAC, UMUCKA-METO, and UMUCKA-UCAM use water vapour climatologies, so no tape recorder diagnostics can be calculated for these models. The model-to-model differences in the amplitude of the anomalies, propagation speed, and attenuation rate are quite large. There are also significant differences in the location of the base of the tape recorder (*i.e.*, the minimum

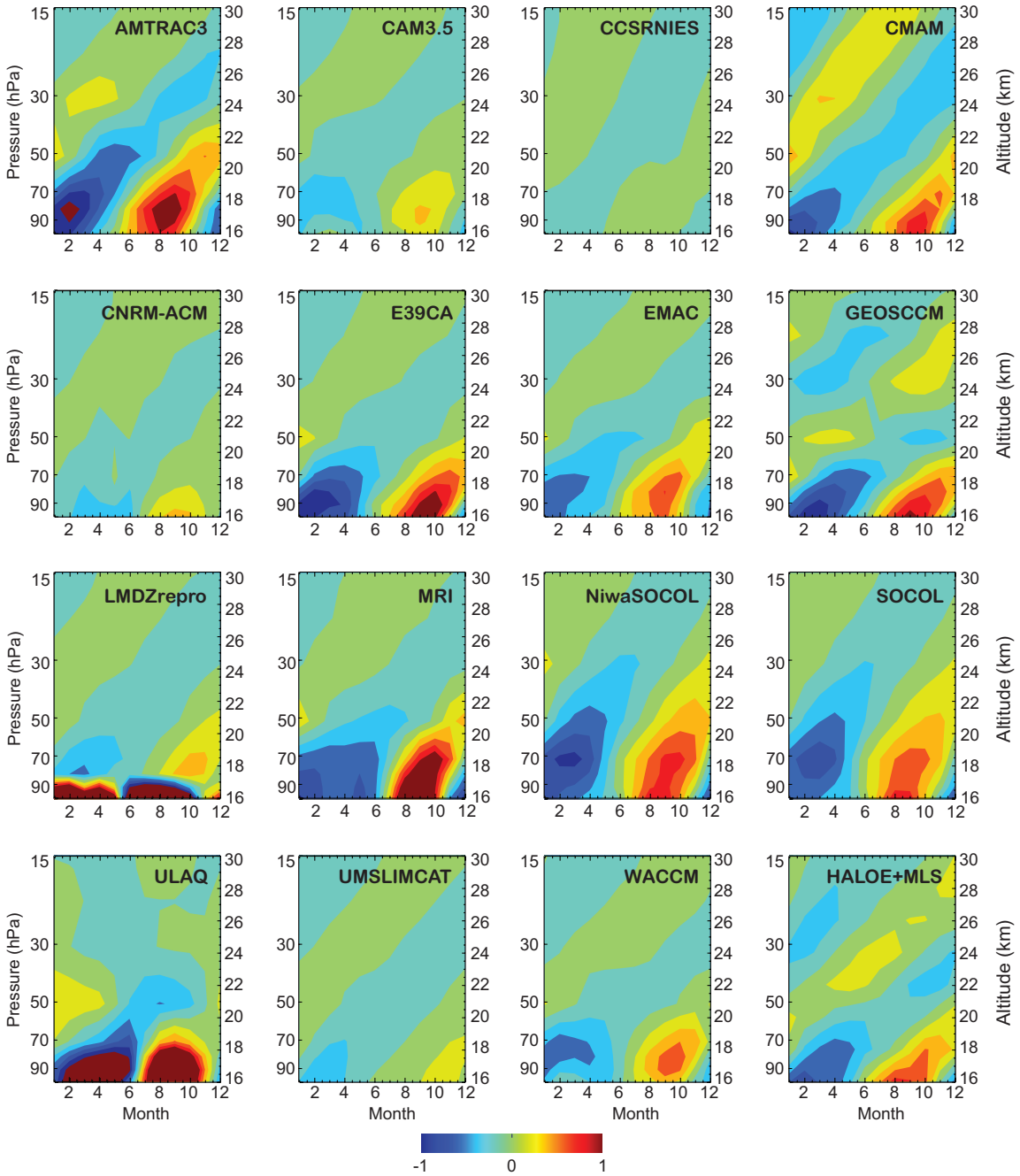

Figure 5.2: Water vapour tape recorder signal from the models and the combined HALOE+MLS data set from Schoeberl et al. (2008). Plots show the deviation in water vapour mixing ratio (ppmv) from the time-mean average profile averaged over 10°N-10°S. The contour interval is the same on all plots and is given at the bottom of the figure.

in water vapour) near the tropical tropopause, but most of the models fall within the range indicated by MLS (3.5 km vertical resolution, tape recorder base at ~80 hPa) and HALOE (2 km vertical resolution, tape recorder base at ~100 hPa). The morphology of the water vapour anomalies is significantly different than observed in several models.

Figure 5.3 compares the phase lag of the tape recorder signal from the models, calculated as the average propagation of the maximum and minimum water vapor anomalies, to the HALOE-derived phase lag (Hall *et al.*, 1999; Eyring *et al.*, 2006). The phase lag is set to zero at the level of maximum amplitude, Z_o, for the observations and for each model, and is plotted as a function of altitude above Z_o. The slope of the phase lag (the phase speed, c) of the tape recorder signal is a measure of the net vertical transport in the tropics (large-scale ascent +

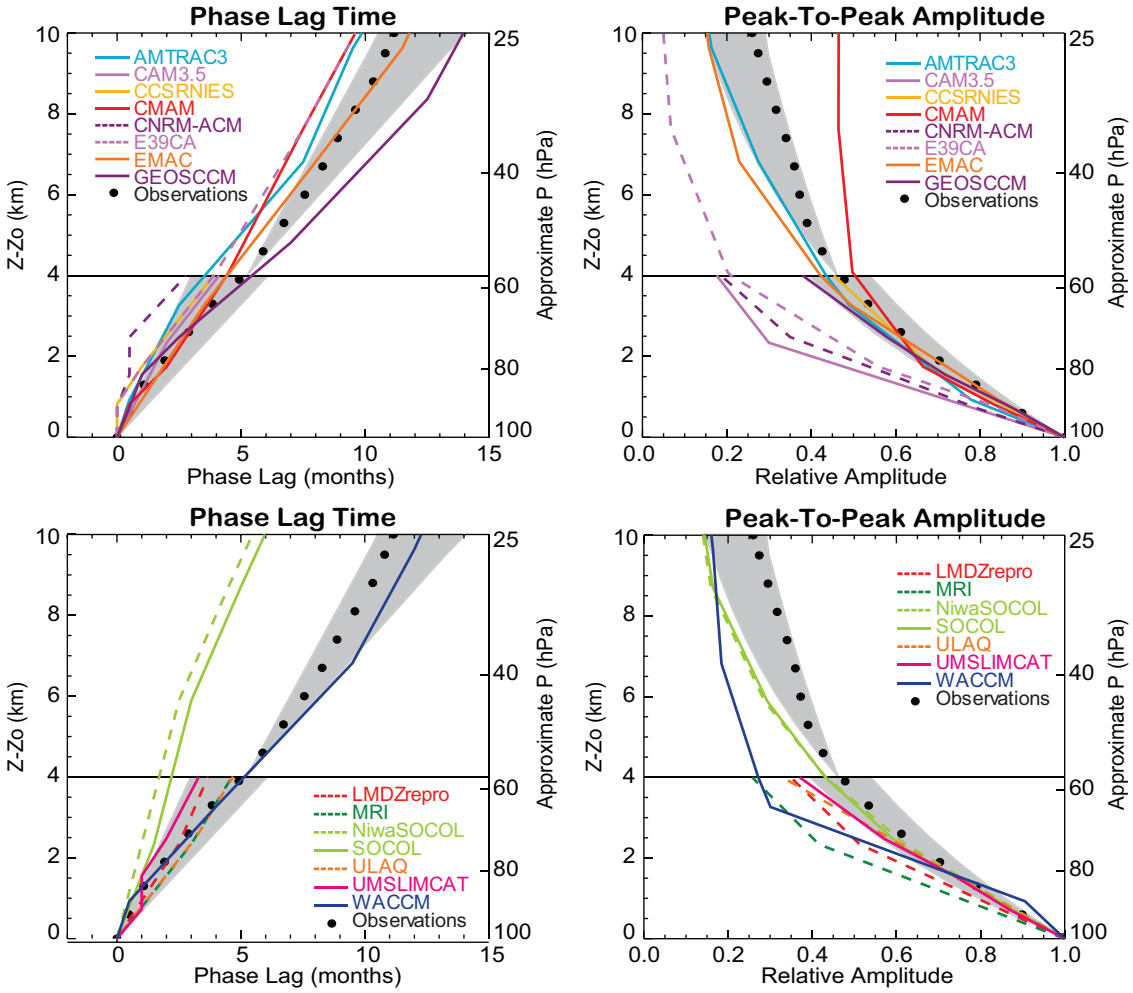

Figure 5.3: Left Panels: Phase lag of the water vapour tape recorder, averaged over 10°S-10°N. The phase lag is set to zero at the level of maximum amplitude, and the vertical coordinate is the distance from that level. The phase lag is the average of the propagation of the maxima and the minima. The models are split into two panels for clarity. Solid circles are HALOE observations. Grey shaded areas indicate the minimum and maximum slopes ± the estimated measurement uncertainty used to evaluate the models in the TLS and TMS. Note that the shaded areas do not represent the uncertainty of the phase lag, but of the phase speed (slope of the phase lag). Right Panels: Amplitude of the tape recorder signal relative to the maximum amplitude as a function of height above the level of maximum amplitude. The grey shaded areas indicate the range of scale heights that can be fit to the observations (± estimated measurement uncertainty).

vertical diffusion) (Mote *et al.*, 1996; Mote *et al.*, 1998; Hall *et al.*, 1999), and the HALOE phase speed has been shown to agree quite well with estimates of the residual vertical velocity in the 16-26 km range (Mote *et al.*, 1998; Schoeberl *et al.*, 2008). The phase speed in the tropical LS (TLS, $Z - Z_o \leq 4$ km) and tropical MS (TMS, $Z - Z_o > 4$ km) is evaluated using a linear fit to the phase lag for both the observations and the models. The phase lag is ill-defined when the amplitude of the water vapour maxima or minima is less than 0.1 ppmv, so it is not analysed if the amplitude falls below this threshold. All of the models have sufficient amplitude for evaluation in the TLS, but only eight do in the TMS. The 0.1 ppmv threshold precludes evaluation of CAM3.5, CCSRNIES, CNRM-ACM, LMDZrepro, MRI, ULAQ, and UMSLIMCAT in the TMS.

Figure 5.4 shows the phase speed calculated from the observations and from the models along the y-axis for the TLS (top panels) and TMS (bottom panels). The mean phase speed derived from a linear fit to the observations is 0.34 mm/s in the TLS and 0.35 mm/s in the TMS. WE08 estimated a standard deviation of 0.05 mm/s for the phase speed derived from the entire altitude range of the observations, but a larger uncertainty is used here since the analysis is split into the TLS and TMS, thus reducing the number of points for obtaining the linear fit. The maximum and minimum phase speeds that can be derived from the obser-

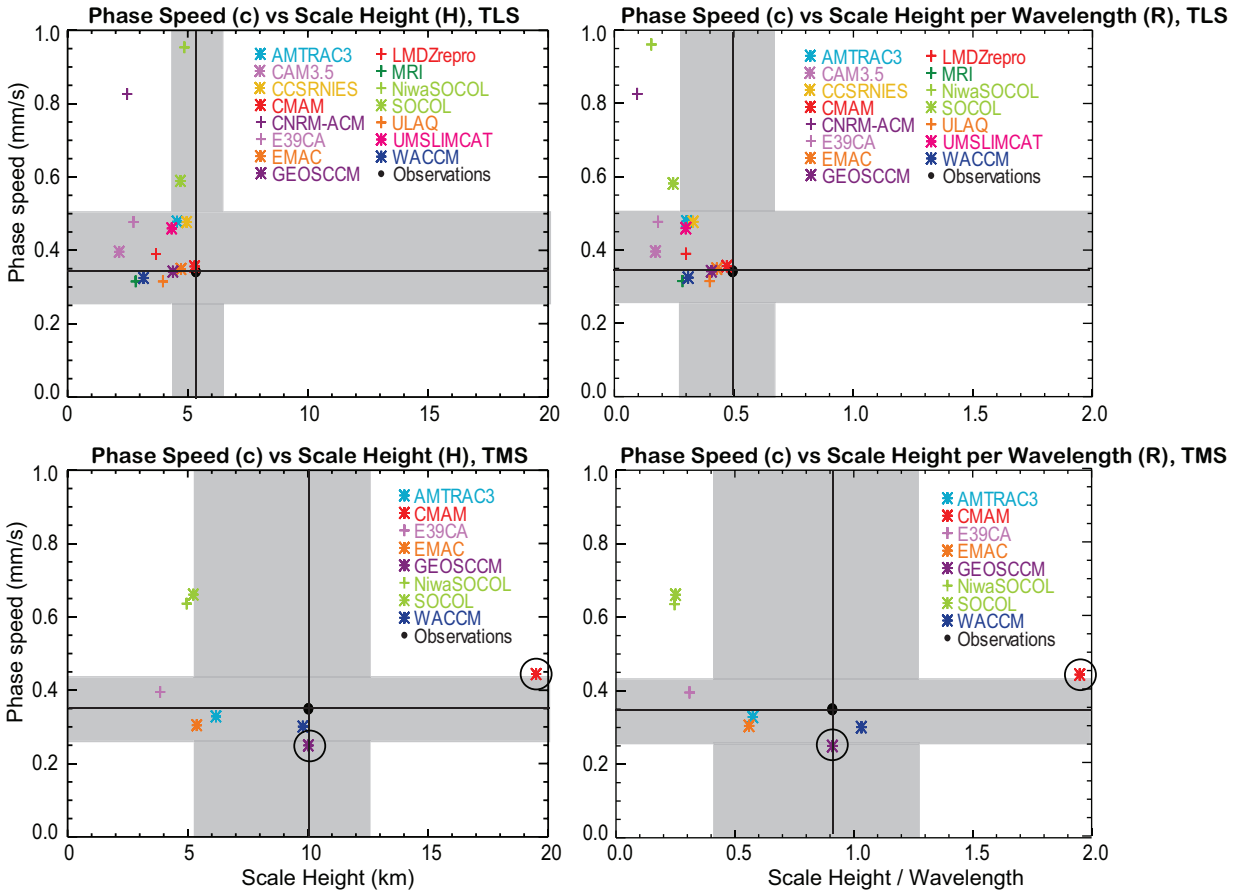

Figure 5.4: Left panels: The tape recorder phase speed, c, calculated as the slope of a linear least squares fit to the phase lag, versus the scale height, H, that satisfies the equation $A=A_0\exp(-z/H)$, where A/A_0 is the relative amplitude of the tape recorder signal. Top panel shows the values for the TLS, bottom panel shows the values for the TMS. The solid black circle shows the observations, and the horizontal and vertical lines show the observed phase speed and scale height, respectively. Horizontal grey shaded regions are bounded by the maximum and minimum slopes that can be fit to the phase lag observations ±1σ measurement uncertainty, and the vertical grey shaded regions are bounded by the maximum and minimum scale heights that can be fit to the observations ±1σ. Right panels: Same as the left panels, except that the scale height per wavelength, R=H/c*1year, is plotted along the x-axis. Vertical grey shaded regions are bounded by H_{max}/c_{min} and H_{min}/c_{max} ±1σ. Note that the faster the phase speed, the larger the difference between R and H. GEOSCCM is circled in the bottom panels because its scale height could not be evaluated in the TMS; the amplitude increases with height due to evaporation of precipitable water. CMAM is circled because its scale height is off the scale of the plot: H=72 km, R=5.2 for CMAM in the TMS.

vations are used as an additional measure of uncertainty, so that the denominator of the WE08 metric becomes $(c_{max} - c_{min}) + 3\sigma$, with the maximum and minimum values of c determined by a linear fit through the observations with the largest least squares residual values in each region, and $\sigma = 0.05$ mm/s as in WE08. The light grey shaded areas in Figures 5.3 and 5.4 show the maximum and minimum slope ± 1σ in each region.

As in CCMVal-1, most of the model phase speeds are greater than observed, indicating too rapid net vertical motion in the tropics. However, the total model spread over the range of observations has decreased since CCMVal-1, as has the difference between modelled and observed phase speeds, suggesting improvements in tropical transport in the models. This will be discussed further in Section 5.5.4. The models that most closely match the observed TLS phase speed are CMAM, EMAC, GEOSCCM, MRI, ULAQ, and WACCM. AMTRAC3, CAM3.5, CCSRNIES, E39CA, LMDZrepro, and UMSLIMCAT all have fast phase speeds relative to the observations, but are within the estimated uncertainties. Three models fall well outside the range of uncertainty in the TLS, with extremely fast phase speeds: CNRM-ACM, NiwaSOCOL, and SOCOL. Of the 8 models evaluated in the TMS, AMTRAC3,

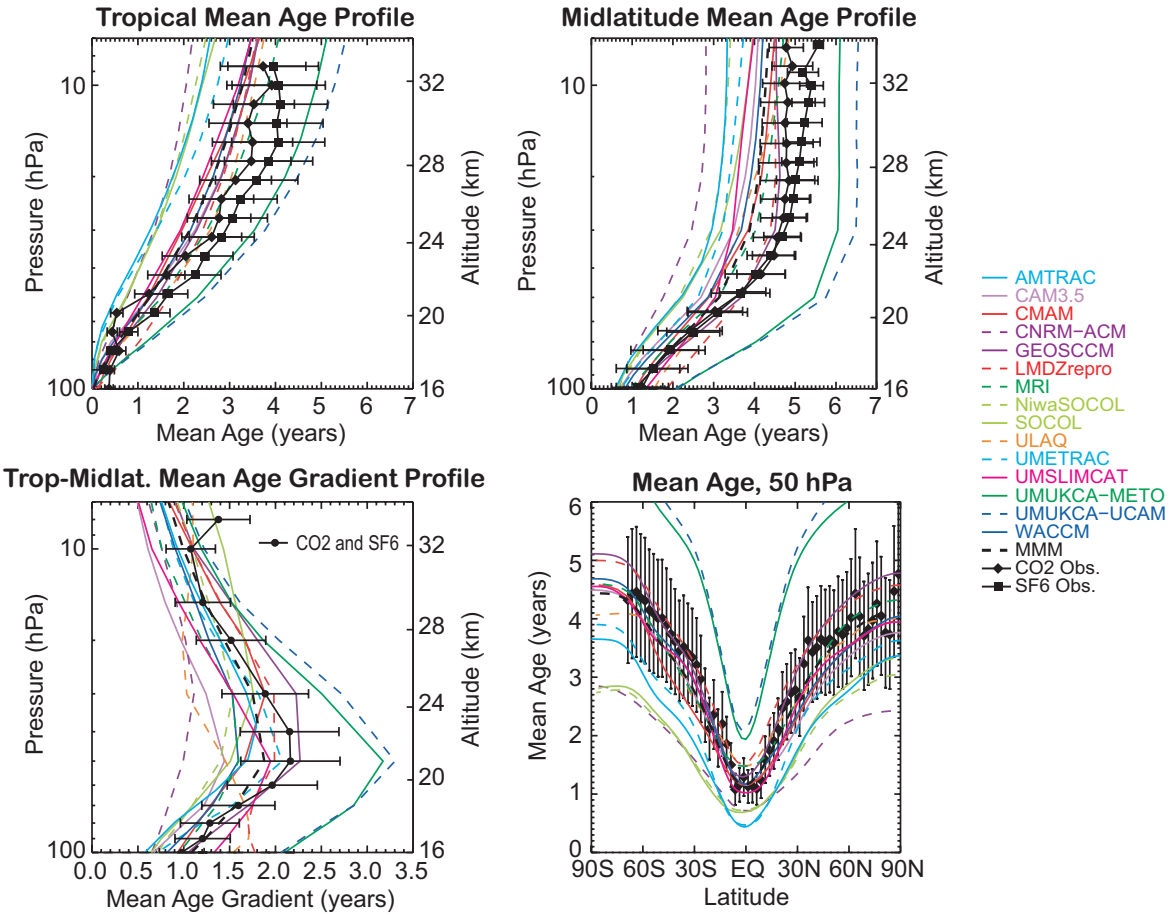

Figure 5.5: Mean age from 15 CCMs and the multi-model mean. Black symbols are the observed mean age profiles derived from CO_2 (diamonds) and SF_6 (squares) for the tropics (10°N -10°S, upper left panel (Andrews et al. (2001)) and mid-latitudes (35°N – 45°N, upper right panel (Engel et al., 2009)), as well as the latitudinal distribution of mean age at 50 hPa (bottom right panel, Andrews et al., (2001)). The bottom left panel shows the difference between the average of the observed tropical profiles and midlatitude profiles on pressure surfaces. All uncertainties shown are 1σ.

E39CA, EMAC, and WACCM fall within the uncertainties of the observed phase speed. CMAM, NiwaSOCOL, and SOCOL have faster than observed TMS phase speeds and the TMS phase speed in GEOSCCM is too slow compared to observations.

5.2.1.2 Ascent from Mean Age Gradients

The mean age of air is the time elapsed since a stratospheric parcel of air was last in contact with the troposphere, and it can be calculated from observations of conserved tracers whose concentrations increase approximately linearly over time. Observations of CO_2 and SF_6 have been used in previous studies to derive empirical estimates of the mean age and to qualitatively evaluate model representations of the residual circulation and mixing (Hall *et al.*, 1999; Eyring *et al.*, 2006).

Neu and Plumb (1999) demonstrated that for a simplified "tropical leaky pipe" model in steady equilibrium, the age difference between the tropics and mid-latitudes along surfaces parallel to the age isopleths depends only on the local tropical vertical velocity and the air mass overhead, assuming that the cross-isopleth diffusion of age is negligible. The age difference is independent of path and of mixing across the edge of the tropics. Any horizontal mass flux from the tropics into mid-latitudes (which acts to decrease the age in mid-latitudes) will be balanced by a decrease in the tropical vertical mass flux (thus increasing the mid-latitude age). In the case of recirculation of air through the tropics, multiple circuits act to age both the tropics and mid-latitudes equally. The result that the age gradient depends only on the ascent rate when diffusion is negligible does not depend on the artificial construct of discontinuities between the tropics and extra-tropics; it applies in general, except in the singular case where the mixing between tropics and mid-latitudes is infinitely fast, as

Table 5.2: CCM Age Tracer Information.

CCM	Tracer Type[1]	Reference Location[2]
AMTRAC3	Stratospheric source	N/A
CAM3.5	Linearly increasing	Equator (0.95°N), 139 hPa
CCSRNIES	Linearly increasing	PBL
CMAM	Stratospheric source	N/A
CNRM-ACM	Linearly increasing	8°S-8°N, below 700 hPa
GEOSCCM	Linearly increasing	Equator at 100 hPa
LMDZ-repro	Linearly increasing	Tropical Thermal Tropopause
NiwaSOCOL	Linearly increasing	PBL
MRI	Linearly increasing	Equator (1.4°S-1.4°N) at 100 hPa
SOCOL	Linearly increasing	PBL
ULAQ	Linearly increasing	15°S-15°N, below 132 hPa
UMSLIMCAT	Pulse source	N/A
UMUKCA-METO	Linearly increasing	PBL
UMUKCA-UCAM	Linearly increasing	PBL
WACCM	Linearly increasing	Equator (0.95°N) at 139 hPa

[1] Linearly increasing - an inert tracer whose concentration grows linearly with time below a given lower boundary; Stratospheric source - direct 'age of air' tracer, where the value of the tracer field in the stratosphere increases by Δt every model Δt; Pulse tracer - given a value of 1.0 during the first month of model simulations and then set to 0.0 afterwards.

[2] Reference location refers to the location used to calculate the mean age fields archived at British Atmospheric Data Centre. For all plots shown the mean age fields were normalised so that mean age = 0 at Equator, 100 hPa.

long as the age difference is taken to be between upwelling and downwelling regions (R. A. Plumb, personal communication). Thus, the difference in mean age between tropics and mid-latitudes can be used to assess tropical ascent independently of quasi-horizontal mixing, which cannot be done using the age itself.

A recent study by Engel *et al*. (2009) used 27 balloon-borne CO_2 and SF_6 profiles measured over 30 years to derive the loss-corrected mean age of air from 35°N-45°N, between 15 and 32 km. Combining these newly available mid-latitude age profiles (**Figure 5.5**, top right panel) with existing tropical profiles from 10°N to 10°S (Figure 5.5, top left panel) (Boering *et al*., 1996; Andrews *et al*., 2001) provides the necessary data to calculate a profile of the tropical-mid-latitude age gradient (Figure 5.5, bottom left panel). The observational uncertainty in the mid-latitude mean ages reflects a combination of the trace gas uncertainties and the variability of mean age over the 30 year period (Engel *et al*., 2009). The fact that the Engel *et al*. (2009) data reflect little or no trend in the Northern Hemisphere (NH) mean age above 24 km is discussed further in Section 5.4. For the purposes of this discussion, the absence of a trend allows for a meaningful comparison between observations collected over the past three decades and present-day model output. The tropical data (Andrews *et al*., 2001) were not reported with uncertainties, but the variability in the published CO_2 and SF_6 profiles suggests that an uncertainty of ±0.5 year for mean ages above 20 km spans the range of observed variability (*e.g*., Eyring *et al*., 2006). Stiller *et al*. (2008) report mean ages derived from global measurements of SF_6 but they cannot be used for this diagnostic because the effects of mesospheric losses were not used in the mean age calculation. Age derived from SF_6 assuming no loss results in mean ages that are up to 1.5 years older than CO_2-derived mean ages in the LS, with much greater differences in the MS.

Table 5.2 describes the type of age tracer used in each model and the reference location used to normalise the mean age where applicable. For the mean age diagnostics, all modelled mean ages were renormalised to 0 on the 100 hPa surface at the equator. The comparison of modelled and observed age difference between the tropics and mid-latitudes is shown in Figure 5.5 (bottom left panel), and the metrics are calculated as an average of the scores at 90, 80, 70, and 50 hPa for the TLS, and as an average of the scores at 30, 20, 15, and 10hPa for the TMS. There is no standard model output level between 50 hPa and 30 hPa, and so the models' ability to capture the maximum age gradient between these levels cannot be assessed.

In agreement with the results for the tape recorder phase speed, most models have smaller than observed age gradients in the TLS, indicating fast ascent. However, as with the tape recorder, most models lie within the observational uncertainty. Only AMTRAC3, CAM3.5, CNRM-ACM, NiwaSOCOL, and SOCOL have age gradients that are smaller than observed and lie outside the observational uncertainty over most of the TLS. Both UMUCKA models, which were not included in the tape recorder analysis, have much larger than observed age gradients (well outside the observational uncertainty), indicating very slow tropical ascent. The age difference profiles in ULAQ

and LMDZrepro have a significantly different shape than observed in the TLS. In the TMS the agreement between the models and the observations is improved, with only CAM3.5, CNRM-ACM, and ULAQ having age gradients smaller than the observational uncertainty over most of the region. Again, the ULAQ age difference profile shape is much different than observed in the TMS. The UMUCKA models are closer to observations in the TMS, but are still the only models with age gradients that are significantly larger than observed.

5.2.1.3 Comparison of Vertical Velocities

The tape recorder phase speed and tropical-mid-latitude age gradient are fully independent measures of the tropical upwelling, and tracer-derived vertical velocities can only be expected to agree with one another when the tracer-dependent terms of the continuity equation are small and the transport circulation closely approximates the residual circulation. Processes that can lead to significant differences in tracer-derived vertical velocities include vertical diffusion, horizontal and vertical eddy tracer fluxes, and rectifier effects between tracer variability and the seasonal or interannual variability of the circulation (see Andrews, *et al*. (1987) for a discussion of the differences between the transport and residual circulation). In models, numerical errors in transport can also lead to significant differences in tracer-derived vertical velocities.

The top left panel of **Figure 5.6** shows a comparison between the tropical vertical velocity calculated from the observed tropical-mid-latitude age gradient (from Figure 5.5, bottom left panel), the vertical velocity calculated by Schoeberl *et al*. (2008) for the combined HALOE-MLS water vapour tape recorder, and the vertical velocity calculated as a simple vertical derivative of the phase lag of the HALOE water vapour observations (from Figure 5.3, left panels). Assuming that age isopleths are roughly parallel to pressure surfaces in the tropics and in the well-mixed portion of the mid-latitudes, the age gradient estimate of the vertical velocity is given by

$$(1+\alpha)H/\alpha\ \Delta\Gamma, \qquad (5.1)$$

where H is a constant scale height (7 km), α is the ratio of the mass of air in the tropics to the mass of air in mid-latitudes, and $\Delta\Gamma$ is the tropical-mid-latitude age difference on pressure surfaces (Neu and Plumb, 1999). The boundaries of the tropical upwelling region used to calculate α are ±25°. The Schoeberl *et al*. (2008) tape recorder vertical velocities were calculated from the phase-lagged correlation coefficient between adjacent levels, and give a much smoother w profile than a simple vertical derivative of the tape recorder phase lag. The three observational tracer-derived estimates of the tropical upwelling show remarkable consistency. Below 26 km, the tape recorder vertical velocity has been shown to agree very well with estimates of the residual circulation (Mote *et al*., 1998; Schoeberl *et al*., 2008). The agreement between the observed age gradient and tape recorder vertical velocities provides further evidence that the transport circulation is a very good approximation to the residual circulation and that the tracer-dependent transport terms are small in the real atmosphere.

The remaining panels of Figure 5.6 show the residual vertical velocity, $\overline{w}^*$, averaged over ±20°, and the age gradient and tape recorder vertical velocities for all of the models. The age gradient velocities are calculated in the same way as for the observations, and the tape recorder velocities are a simple vertical derivative of the phase lag from Figure 5.3. The tape recorder velocities are plotted starting at the level of maximum amplitude for each model tape recorder rather than normalised to a common height as in Figure 5.3. The solid black line in each panel is the mean of the observed age gradient vertical velocity and the Schoeberl *et al*. (2008) tape recorder vertical velocity from ~90-25 hPa. Above and below that region the age gradient vertical velocity is used. The grey shaded region shows the corresponding uncertainties.

Of the ten models that have all three vertical velocity estimates, there is relatively good agreement between the three different upwelling calculations and good agreement with the observations for CMAM, GEOSCCM, MRI (which only has tape recorder results in the TLS), and WACCM. The differences between the tracer-derived vertical velocities and $\overline{w}^*$ are relatively small for LMDZrepro and ULAQ, but as noted in Section 5.2.1.2, both show age gradient profiles that are significantly different than observed (though LMDZrepro agrees well with observations in the TMS). CNRM-ACM has very fast tracer-derived vertical velocities, which differ significantly from $\overline{w}^*$. CAM3.5 shows significant differences between the age gradient vertical velocity, which is considerably faster than observed, and the tape recorder and residual circulation vertical velocities, which compare well to the observations. SOCOL and NiwaSOCOL also have large differences between the tracer-derived velocities; the tape recorder velocity is much faster than observed, while the age gradient velocity is just outside the range of observational uncertainty in the TLS and compares fairly well to the observations in the TMS. Despite very similar tracer-derived vertical velocities, SOCOL and NiwaSOCOL have very different $\overline{w}^*$ profiles, with SOCOL's $\overline{w}^*$ agreeing very well with the observed tracer velocities and NiwaSOCOL's $\overline{w}^*$ being much faster than any other model's; see also Figure 4.9. EMAC and E39CA do not have an age tracer. EMAC shows very good agreement between its tape recorder vertical velocity and $\overline{w}^*$ as well as very good agreement with the observations. The E39CA tape recorder has a layer of infinite phase speed (zero phase lag) above the tropopause, but shows rel-

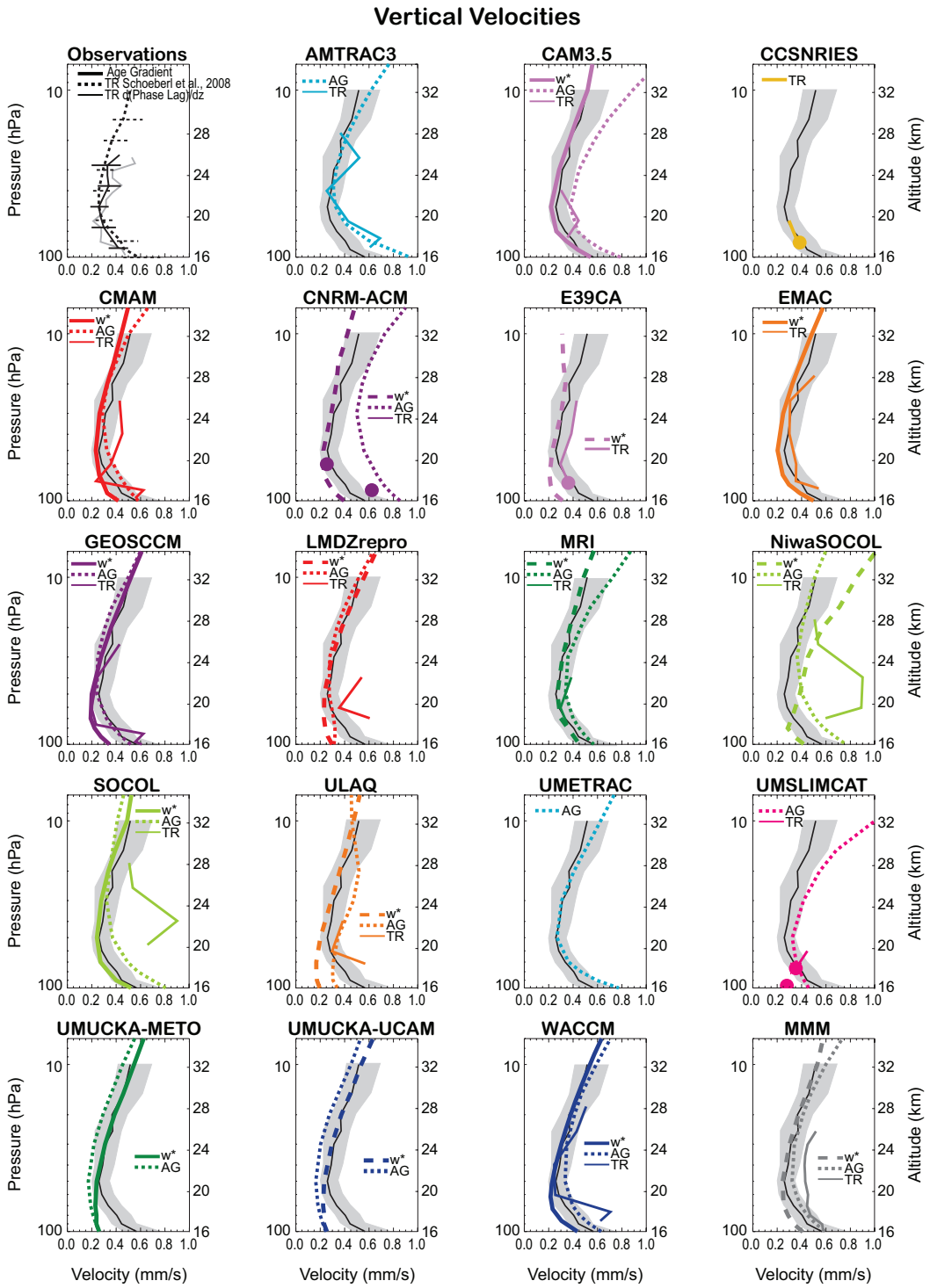

Figure 5.6: Comparison of the tropical vertical velocities derived from the tape recorder (TR) and mean age gradient (AG), as well as model residual vertical velocities, $\bar{w}^*$. The top left panel shows w calculated from the observed AG with uncertainties, w calculated by Schoeberl et al. (2008) for the combined HALOE-MLS TR with uncertainties, and w calculated as a simple vertical derivative of the phase lag of the HALOE TR. The remaining panels show $\bar{w}^*$, averaged over ±20°, and the AG and TR w's where available for all of the models. Tape recorder w's are represented by large dots where the adjacent levels give infinite phase speeds. The solid black lines and grey shaded regions in the model panels are an average between the observed AG and Schoeberl et al. (2008) TR w's and uncertainties between 90 and 25 hPa. Above and below those levels, the observed AG w and uncertainties are used.

atively good agreement with the observations and with $\overline{w}^*$ above 80 hPa. UMUCKA-METO and UMUCKA-UCAM do not have tape recorder signals, but show relatively close agreement between their age gradient vertical velocities and $\overline{w}^*$, both of which are very slow in the TLS relative to the observations. The residual vertical velocity could not be obtained for AMTRAC3, CCSRNIES, UMETRAC, and UMSLIMCAT. The tape recorder and age gradient vertical velocities agree very well in AMTRAC3, and are very fast compared to the observations in the TLS. The age gradient vertical velocity is also too fast in UMETRAC just above the tropopause, but then agrees very well with observations throughout the rest of the domain. UMSLIMCAT has a layer of infinite phase speed, but overall there is reasonable agreement between the tape recorder, age gradient, and observed velocities in the TLS. In the TMS the age gradient vertical velocity in UMSLIMCAT is much faster than observed. The only vertical velocity available from CCSRNIES is from the TLS tape recorder, which has a layer of infinite phase speed below 90 hPa but then agrees well with observations.

The multi-model mean (MMM) residual vertical velocity and age gradient vertical velocity are fairly consistent with each other and agree closely with observations. The MMM tape recorder vertical velocity was calculated using only the 8 models whose tape recorders could be analysed throughout both the TLS and TMS (the infinite phase speed at the lowest level in E39CA is not included in the mean). It reflects the very fast tape recorder velocities of SOCOL and NiwaSOCOL, and is significantly faster than observations and than the other models' vertical velocities. Overall, the models present a consistent, coherent picture with respect to the transport circulation: when the tracer-derived vertical velocities agree with one another, they also agree with the observed residual vertical velocity; when there are differences between the tracer-derived vertical velocities, they also differ substantially from the residual circulation, indicating a substantial role for tracer-dependent terms in the transport circulation.

5.2.2. Tropical-Midlatitude Mixing

5.2.2.1 Tape Recorder Amplitude

The tape recorder amplitude decays with height due to both vertical diffusion and dilution by mid-latitude air. Vertical diffusion plays a modest role in the decay from 19-24 km, where ascent rates are slow and the tropics are relatively isolated, but dilution accounts for most of the attenuation of the observed signal (Mote *et al*., 1998; Hall *et al*., 1999). However, Hall *et al*. (1999) showed that diffusion can play a significant role in the attenuation of the water vapour tape recorder in models. Three of the models shown here (CNRM-ACM, SOCOL, and NiwaSOCOL) have tape recorder phase speeds that are considerably faster than $\overline{w}^*$, indicating that they may be in the "high diffusion" regime described by Hall *et al*. (1999), in which case diffusion can account for a large portion of the amplitude attenuation. Nevertheless, most of the models have $c \approx \overline{w}^*$, indicating that they are in the "low diffusion" regime and thus the attenuation is a measure of dilution by mid-latitude air.

Figure 5.3 shows the vertical profile of the decrease in the peak-to-peak amplitude of the tape recorder water vapour anomalies relative to the maximum peak-to-peak amplitude for the models and for HALOE observations. For a given dilution profile, rapid ascent will result in less attenuation of the signal than slower ascent. To isolate the effect of mixing between the tropics and mid-latitudes as much as possible, the influence of the phase speed on the amplitude attenuation is removed by evaluating the scale height of the amplitude decay, H, relative to the vertical wavelength (λ=c*1 year), so that the metric is R=H/λ, which provides a better measure of the dilution rate than H itself. As in WE08, H is determined by an exponential fit to the HALOE observations and the models, with the relative amplitude A/A_0 described by exp(-z/H). As with the phase speed, separate fits are applied in the TLS and TMS and the maximum and minimum values of R are used as an additional measure of uncertainty so that the denominator of the WE08 metric becomes $(R_{max} - R_{min}) + 3\sigma$. R_{max} is given by H_{max}/C_{min}, where $H_{\max}$ is the maximum scale height that can be fit to the observations (*i.e*., the exponential passes through the largest residual of the fit in each region), and $c_{\min}$ is as defined in Section 5.2.1.1. R_{min} is likewise equal to H_{min}/C_{max}. An observational uncertainty of 20% is estimated for R, similar to the value used by WE08.

Figure 5.4 shows H (left panels) and R (right panels) for the observations and for the models. The observational values of H are 5.3 km in the TLS and 10.0 km in the TMS, corresponding to R values of 0.5 and 0.9, respectively. These values are in good agreement with Hall *et al*. (1999). The light grey shaded areas in Figures 5.3 and in the left panels of 5.4 show H_{max} and $H_{min} \pm 1\sigma$ in each region (with σ estimated as 20%, as for R). The shaded regions in the right panels of Figure 5.4 show R_{max} and $R_{min} \pm 1\sigma$. R was not calculated for the CCMVal-1 models in Eyring *et al*. (2006), and it is difficult to compare the differences in the amplitude attenuation given the differences in phase speed. However, most CCMVal-1 models attenuated the tape recorder signal too strongly despite all having fast phase speeds, indicating too much dilution by mid-latitude air and/or too much vertical diffusion. Here also, most of the models attenuate the signal too quickly in the TLS compared to the observations ($R < R_{obs}$). CMAM, EMAC, GEOSCCM, and ULAQ are the best performing models. All other models have values of R near or outside of the range of uncertainty in the measurements. For mod-

els with $c \approx \overline{w}^*$, this indicates too much mixing across the subtropics. In the TMS, WACCM is the best performing model. CMAM has extremely isolated ascent relative the observations, with an *R* value of 5.2. All of the other models have *R* values less than the observations, ranging from ~0.6 (AMTRAC3 and EMAC) to less than 0.35 (E39CA, NiwaSOCOL, and SOCOL), again indicating too much tropical-extra-tropical mixing and/or vertical diffusion for this limited set of models. The GEOSCCM tape recorder amplitude was not evaluated in the TMS because it increases with height over part of this region. GEOSCCM carries precipitable water too high into the stratosphere, and the increase in amplitude results from re-evaporation of condensed water.

5.2.2.2 Tropical CH_4 Vertical Gradient

Methane is long-lived in the lower and middle stratosphere but is destroyed by $O(^1D)$, OH and Cl radicals in the upper stratosphere; its annual-average photochemical lifetime is one year at 3 hPa and increases rapidly with decreasing altitude. Thus, below 3 hPa there is almost no photochemical loss and the profile's vertical gradient is primarily controlled by the balance between ascent and quasi-horizontal mixing across the subtropics. This diagnostic tests the ability of the models to represent the observed balance between these two processes, but, like the tape recorder amplitude, it is sensitive to vertical diffusion. The models are evaluated in two seasons using 9 years of HALOE CH_4 from 10°S-10°N. At the lowest level of the HALOE observations (68 hPa), 10 years of seasonal mean model output are normalised to the 9-year tropical mean CH_4 from HALOE. Model output is interpolated to the 9 HALOE vertical levels (68-3 hPa, or ~18-42 km with 3 km spacing) and the gradients are evaluated over three regions: TLS (68-31 hPa), TMS (31-10 hPa), and tropical upper stratosphere (TUS) (10-3 hPa). The ratio of the mixing ratios at the top and bottom of each altitude range is evaluated and compared to the observations. The HALOE uncertainties are calculated from deviations from the seasonal mean gradients, averaged over 9 years of observations, and include the effects of QBO and interannual variability.

The mixing ratio vertical gradient is evaluated rather than the mixing ratio itself so that poor transport at one level will not affect the score of the levels above it. For example, if a model has too much subtropical mixing and/or ascent in the TLS, but the right amount above, the profile in the TMS will be offset from the HALOE profile but it will have the same vertical gradient. Each model's grade is the average of the score for the gradients in the TLS, TMS, and TUS, for two seasons, January, February, March (JFM) and July, August, September (JAS).

Although the quantitative assessment is based on the gradients, for consistency with previous work (Eyring *et al*., 2006) **Figure 5.7** shows HALOE and model CH_4 tropical seasonal mean profiles. Many models reproduce the profiles fairly well. The bottom panel of Figure 5.7 shows that the models perform about equally well in the TLS and TMS. The spread of model performance is greater during JAS than during JFM, especially in the middle and upper stratosphere. A direct comparison to CCMVal-1 is difficult because Eyring *et al*. (2006) showed CH_4 profiles for March. However, it appears that the spread in model performance has decreased. Approximately the same percentage of the models show good agreement with the observations in the two assessments, but the worst-performing models are much closer to the observations in CCMVal-2 than in CCMVal-1.

It is possible to infer information about mixing in the models and compare it to the results from the tape recorder attenuation by using the ascent rate information from Section 5.2.1 and accounting for the tendency for rapid (slow) vertical transport to decrease (increase) the vertical gradient of CH_4 for a given tropical-mid-latitude mixing profile. The TLS and TMS definitions used here (~18-24 km and ~24-32 km) differ somewhat from the altitude ranges used in the tape recorder attenuation analysis (~17-21 km and 21-27 km, respectively, depending on the location of the water vapour minimum), but in most cases the overall behaviour of the CH_4 profile in each region does not depend strongly on the exact altitudes used and there is consistency between the two diagnostics. CNRM-ACM, SOCOL, and NiwaSOCOL match the observed CH_4 gradients well. This is not necessarily inconsistent with their very strong attenuation of the tape recorder signal, given that they may have significant vertical diffusion. However, the comparison between mixing diagnostics is difficult, both because of the direct impact of vertical mixing on tracers and because diffusion generates inconsistencies in the estimates of vertical velocity. Of the remaining 12 models that have output for both diagnostics in the TLS, all but two (EMAC and ULAQ) show qualitative agreement between the tape recorder and CH_4 mixing diagnostics. In both EMAC and ULAQ, the TLS CH_4 gradients are much stronger than observed, indicating too much mixing, while the tape recorder signal closely matches the observations. The differences in the diagnostics do not appear to be related to the differences in TLS altitudes. Three models (AMTRAC3, CMAM, and E39CA) show significant differences between the mixing inferred from the CH_4 gradients and that inferred from the tape recorder signal in the TMS. The tape recorder analysis implies too much mixing for AMTRAC3 and E39CA and almost no mixing for CMAM. The CH_4 gradients, on the other hand, indicate that CMAM and E39CA have good mixing and AMTRAC3 has too little. For AMTRAC3 and CMAM, which have the output available for the tropical age and N_2O PDF mixing diagnostics, the results of the CH_4 analysis are supported by the

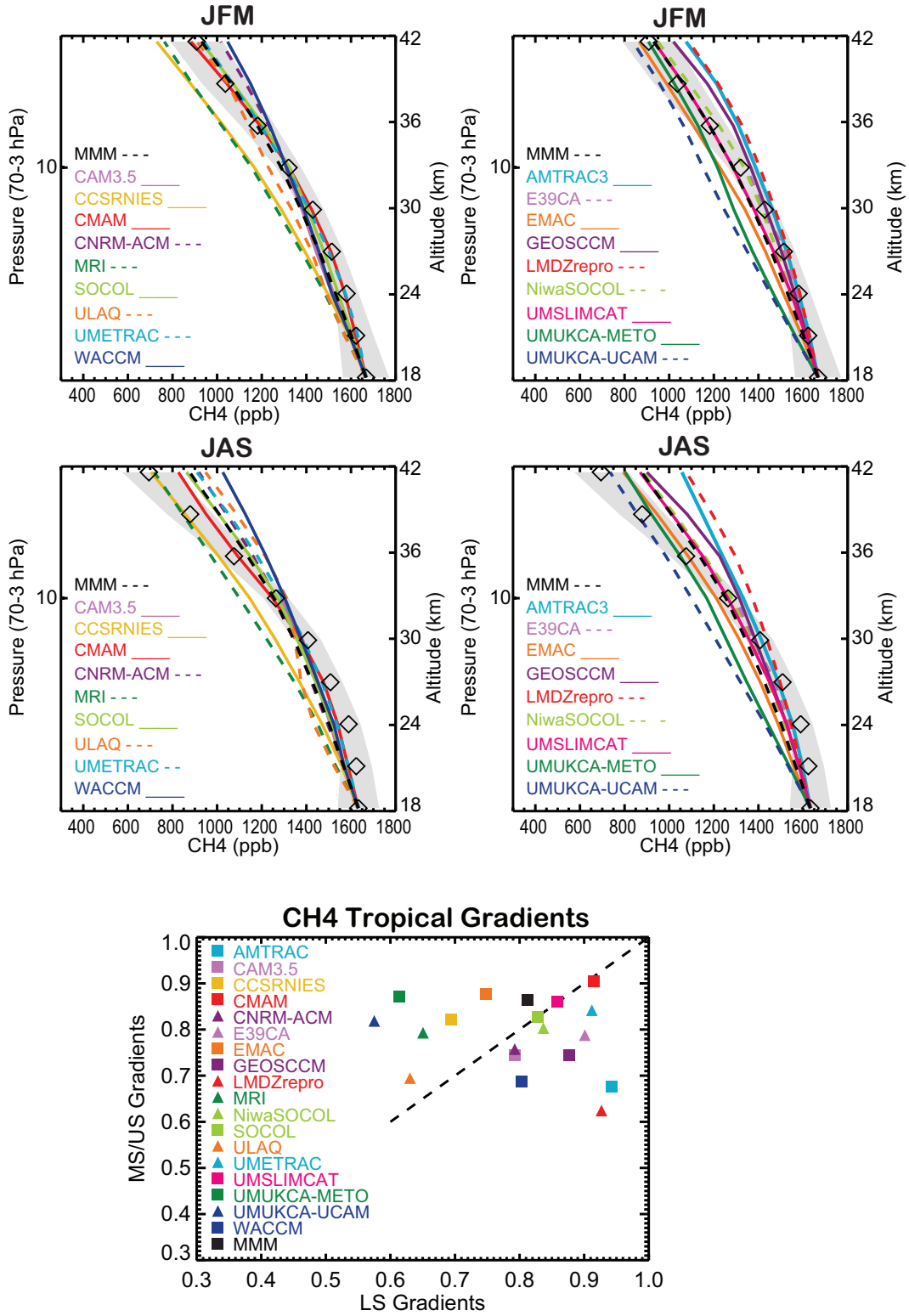

Figure 5.7: Tropical (10°N-10°S) CH_4 profiles from all CCMs in two seasons compared to HALOE mean profiles. Diamonds show the HALOE mean CH_4 profile for Jan-Feb-Mar (top panels) and Jul-Aug-Sep (bottom panels). The black dashed line is the profile from the multi-model mean (MMM). The grey shading shows the 1σ range of the observations. The bottom panel compares the model scores for the gradient of the CH_4 profile in the LS versus the scores in the MS.

other mixing diagnostics.

Three models did not have the output for the tape recorder analysis (UMETRAC, UMUCKA-METO, and UMUCKA-UCAM) and eight others have tape recorder output but could not be evaluated in the TMS (CAM3.5, CCSRNIES, CNRM-ACM, GEOSCCM, LMDZrepro, MRI, ULAQ, and UMSLIMCAT). CNRM-ACM is discussed above. UMETRAC was found to have very good ascent using the age gradient diagnostic, and its CH_4 gradients are good in both the TLS and TMS, indicating good mixing across the subtropics in both regions. The CH_4 profiles in the UMUCKA models fall off very rapidly in the TLS, consistent with their very slow ascent. No additional information can be inferred about their tropical-extra-tropical mixing. In the TMS their CH_4 profiles show much better agreement with observations, indicating too little mixing given their slow circulations. GEOSCCM, LMDZrepro, and MRI have good TMS tracer-derived ascent rates. The TMS CH_4 gradients are too weak in GEOSCCM and LMDZrepro, indicating too little mixing, and too strong in MRI, indicating too much mixing. ULAQ has relatively fast age gradient ascent in the TMS. Its CH_4 profile matches the observations in JFM, suggesting too much mixing, but the gradient is much weaker in JAS, which is consistent with the fast ascent. CAM3.5 and UMSLIMCAT have very fast TMS age gradient ascent rates. The CH_4 gradients are very weak in CAM3.5, while UMSLIMCAT matches the observations fairly well. This suggests too much tropical-extra-tropical mixing (particularly in CAM3.5) given the rapid tropical ascent. However, while the residual vertical velocity was not available for UMSLIMCAT, the difference between the age gradient ascent and $\overline{w}^*$ is large in CAM3.5, which points to the possible importance of vertical diffusion. CCSRNIES has no TMS ascent diagnostics. Its CH_4 profiles fall off too quickly with height in the TMS, indicating too much mixing across the subtropics unless the ascent is very slow. The MMM CH_4 gradient is slightly strong in the TLS and slightly weak in the TMS but shows overall good agreement with the observations, suggesting only small net biases in tropical-extra-tropical mixing.

5.2.2.3 Tropical Mean Age

The top left panel of Figure 5.5 shows the tropical (10°S-10°N) annual mean age profile for the CCMs, plotted with the mean age derived from CO_2 (black squares) and SF_6 (black diamonds with ±25% error bars) (Andrews *et al*., 2001). The tropical mean age reflects the combined effects of large-scale ascent, vertical diffusion, and horizontal mixing across the subtropics. In the TLS, the mean age is a relatively local diagnostic of the circulation and mixing, but understanding a model's age becomes more complicated higher up in the tropics, where the integrated effects of transport below make it more difficult to diagnose reasons for deviations from the observations. Furthermore, the influence of in-mixing of mid-latitude air varies with height since the age difference increases in the TLS and decreases in the TMS.

The models are evaluated at four pressure levels in the TLS (90, 80, 70, 50 hPa) and TMS (30, 20, 15, 10 hPa), and the score for each region is the average of the scores and individual pressure levels. Most of the models have younger mean ages than the observations; they do not reproduce the rapid increase in mean age from 60 to 30 hPa, which causes an offset in the profiles throughout the TMS. Seven models fall outside the range of observational uncertainty (AMTRAC3, UMETRAC, CNRM-ACM, NiwaSOCOL, SOCOL, UMUCKA-METO, and UMUCKA-UCAM). The UMUCKA models have older than observed air throughout the tropical stratosphere, reflecting their very slow tracer ascent rates. However, they match the shape of the observed profile from 60 to 15 hPa, suggesting, in agreement with the CH_4 profiles, that their mean ages in this region reflect a balance between slow ascent and very little mixing. The remaining five models have very young mean ages, and all except UMETRAC have very fast tracer-derived vertical velocities over some or all of the altitude range considered. Furthermore, there are indications that three of these models (CNRM-ACM, NiwaSOCOL, and SOCOL) may have excessive vertical diffusion. AMTRAC3's young mean ages in the TLS are consistent with its apparently rapid ascent, but it becomes progressively younger than the observations all the way up to 20 hP despite very good ascent rates in the TMS. The CH_4 gradients from AMTRAC3 suggest, in disagreement with the tape recorder analysis, that there is too little in-mixing from mid-latitudes in the TMS. The age profile supports that conclusion. UMETRAC, which diverges from AMTRAC3 in the TMS and maintains a relatively constant offset from the observations, has CH_4 gradients that are much closer to observed than AMTRAC3.

Of the eight models that generally fall within the range of observational uncertainty, CAM3.5 and UMSLIMCAT have relatively good tropical mean ages that must reflect a balance between rapid net vertical transport and excessive tropical-extra-tropical mixing. LMDZrepro and ULAQ have fairly complicated ascent and mixing profiles and the tropical ages do not provide any clear indication of mixing. CMAM, GEOSCCM, MRI, and WACCM have both good tropical mean age and good ascent rates. GEOSCCM and WACCM have too little tropical-extra-tropical mixing in the TMS, which likely contributes to their slightly young ages. As with AMTRAC3, the CMAM age profile shows agreement with the CH_4 gradients rather than the tape recorder analysis. The complete lack of TMS mixing implied by the tape recorder attenuation would likely yield considerably younger ages than seen here. MRI has the best tropical mean age profile despite diagnostics that indicate that it

has significantly greater than observed tropical-extra-tropical mixing. However, the fact that MRI has older ages than any of the other models with good ascent is consistent with it having more mixing than the other models. The MMM tropical mean age profile closely matches the profile of this cluster of models with good performance; the models with poor performance largely cancel each other out.

5.2.2.4 Tropical-Midlatitude N_2O PDFs

N_2O is a long-lived tracer that decreases with height in the stratosphere. Its distribution is controlled by a balance between the large-scale circulation, which acts to steepen its isopleths, and stirring by wave activity, which acts to flatten its isopleths in the stirring region but produces very strong gradients at its edges. Probability distribution functions of satellite measurements of long-lived tracers such as N_2O show multiple modes, with three modes in the winter hemisphere corresponding to the tropics, the well-mixed surf zone, and the polar vortex, and two modes in the summer hemisphere corresponding to the tropics and the extra-tropics (Sparling, 2000; Neu *et al*., 2003). The minima between the modes correspond to the strong tracer gradients marking the transitions between tropical and extra-tropical air and between the mid-latitudes and the polar vortex. PDFs of N_2O have been used to assess the ability of models to reproduce tropical isolation in the middle and upper stratosphere (Douglass *et al*., 1999; Strahan and Douglass, 2004; Gray and Russell, 1999). Douglass *et al*. (1999) used CLAES data on isentropic surfaces between 10°S and 45°N to construct N_2O PDFs to benchmark the performance of the GMI-CTM. An inability to maintain the separation and the depth of the minimum between the tropical and mid-latitude modes indicates that a model has too much tropical-extra-tropical mixing, which acts to homogenize the distribution.

Recent MIPAS and MLS observations of N_2O are used to determine the ability of the models to maintain the correct tropical isolation. The MIPAS N_2O data are a product of the Institute of Meteorology and Climate Research in Karlsruhe (IMK), updated from Glatthor *et al*. (2005) for the period July 2002 to March 2004. The data has a small positive bias below 25 km (Gabriele Stiller, personal communication). The MLS data cover the years 2004 to present (Lambert *et al*., 2007). Because the observational periods are relatively short and do not overlap, there are differences between the two data sets that are related primarily to QBO variability. By combining the data and using the differences between the observations to define the uncertainty, the influence of the QBO on the analysis is minimised. The models are evaluated at three isentropic levels (600 K, 800 K, and 1000 K; ~24-34 km) in each hemisphere using seasonally-averaged PDFs at the locations and seasons that MIPAS and MLS both indicate a bimodal distribution. The quantitative assessment of the PDFs is based on how well they reproduce the observed separation and relative amplitudes of the modes of the distribution.

Observations and model output are processed in the same way. Instantaneous profiles are binned according to season and isentropic level for each hemisphere (NH: 10°S-45°N; SH: 45°S-10°N). All available data for the observational periods of each instrument are used. The model analyses use instantaneous data from the most recent 4-5 years of the REF-B1 integrations. The characteristics of the PDFs do not tend to be very sensitive to the chosen period. To remove any bias in the N_2O mixing ratio, the PDFs are re-scaled to give a mean value of zero for the distribution. The relative maximum on each side of zero is determined. To determine whether the maxima are well separated, the sensitivity of the location of the maxima to small offsets with respect to zero are tested. If the same maxima are detected, it is assumed that the maxima are well separated, and the relative amplitude (ratio of the right hand peak over left hand peak amplitude) and peak-to-peak separation (difference between the N_2O values for the right peak minus the left peak) are calculated.

The uncertainty is estimated by calculating the relative difference between MIPAS and MLS, so that the metric is

$$1 - \frac{|X_{mod} - X_{obs}|}{n|X_{MIPAS} - X_{MLS}|} \quad , \qquad (5.2)$$

with $X_{obs}=(X_{MIPAS}+X_{MLS})/2$, for each hemisphere at each isentropic level during each season. The separation is generally easier to capture than the relative amplitude and can be tested more stringently, so n=1 is used for the separation and n=2 for the relative amplitude. If the model does not have a bimodal PDF, it scores zero. The final grade of the model is averaged over both hemispheres for all of the isentropic levels and seasons in which MIPAS and MLS both show unambiguously bimodal distributions (12 out of a possible 24).

Figure 5.8 shows the N_2O distribution as a function of potential temperature in the NH spring for the MIPAS and MLS observations and the 16 models that submitted the necessary output for this diagnostic. While this figure illustrates model behaviour for only one season and hemisphere, it provides a good example of model performance. CAM3.5 and ULAQ do not show a good barrier to subtropical mixing. All other models have clearly bimodal distributions at most levels 600 K-1000 K. However, CCSRNIES, EMAC, and UMSLIMCAT do not show enough separation between the tropical and mid-latitude peaks at many levels, while GEOSCCM, UMUCKA-UCAM, UMUCKA-METO, and WACCM show too much separation. **Figure 5.9** shows the PDFs for NH and SH spring at a single level (800K, ~10 hPa). While there are significant differences between the NH and SH in both the

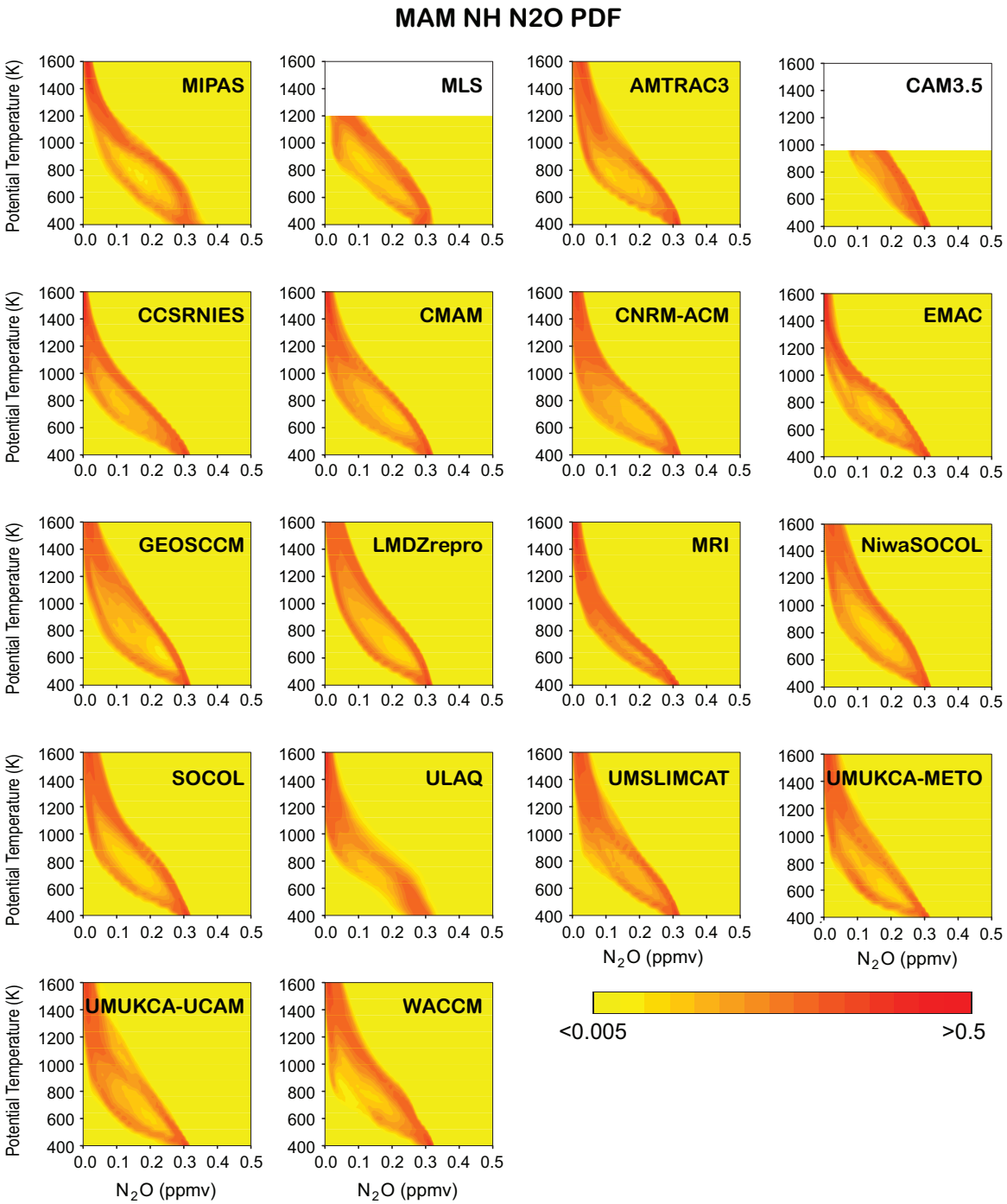

Figure 5.8: Contoured probability distribution functions of N_2O for 10°S-45°N for NH spring (March-April-May) for 16 CCMs. MIPAS and MLS observations are shown in the first two panels. The contour interval is the same for all plots, and is shown in the lower right portion of the figure. Deep orange shows the mixing ratio of the most probable N_2O values. Both MIPAS and MLS show a deep minimum (yellow) in their bimodal PDFs for levels 600 K-1000 K (~24-34 km). This indicates a strong barrier to tropical-mid-latitude mixing in the NH during this season.

observations and the models, the model behaviour is generally consistent in the two hemispheres. Exceptions are CAM3.5, which does not have a bimodal distribution in the NH but shows a clearly bimodal distribution in the SH (though the peaks are not well-separated); LMDZrepro, which shows good separation between the peaks in the NH but too much separation in the SH; and WACCM, which shows too much separation between the peaks in the NH

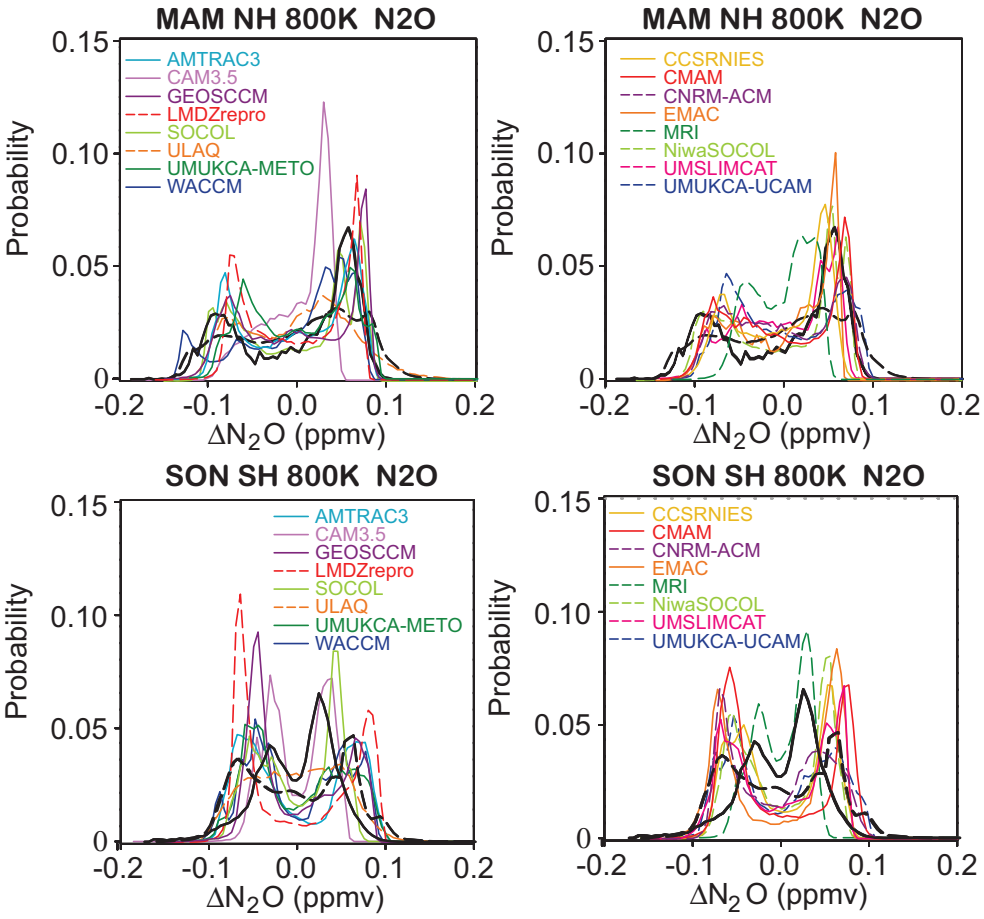

Figure 5.9: Probability distribution functions of N_2O on the 800 K surface for NH spring (top panels) and SH spring (bottom panels). The models are split into two panels for clarity. For the observations and each CCM, N_2O values have been scaled so that the mean value of the distribution is zero. MIPAS observations are black solid lines, MLS observations are black dashed lines. MIPAS and MLS data are from different years. Although they do not show the same separation, both sets of observations are clearly bimodal at these levels in these seasons.

but very good agreement with observations in the SH. On average, AMTRAC3, CMAM, GEOSCCM, LMDZrepro, SOCOL, NiwaSOCOL, and WACCM show very good performance and CAM3.5, MRI, and ULAQ show the poorest performance on this diagnostic. The results for SOCOL and NiwaSOCOL are somewhat surprising given that these two models were shown to have serious transport problems below ~26 km. However, the N_2O PDFs are evaluated from ~24-34 km, where these models show much better agreement with other observations.

5.3 Transport Diagnostics for the Extra-tropics

Trace gas composition in the extra-tropics is affected by many seasonally varying processes. At 70 hPa, the net vertical motion of the annually-averaged Brewer-Dobson circulation equatorward of 40° is upward while poleward of 40° it is downward (Rosenlof, 1995). Young air ascending in the tropics is exported to the mid-latitudes by extra-tropical planetary wave activity, which varies in strength with height and season. In summer, mixing between the mid- and high latitudes is weak because stirring by planetary wave activity is at a minimum. From late fall through early spring, the polar vortex forms, creating a barrier to transport between the mid- and high latitudes. Inside the vortex there is strong, largely unmixed descent, particularly in the Antarctic. In the mid-latitudes, meridional gradients are weak as a result of mixing by strong planetary wave activity (the 'surf zone'), particularly in the NH. The extra-tropical diagnostics presented here cannot isolate and evaluate the effects of a single process, rather, they evaluate the net effect of multiple transport processes. Taken together, these diagnostics evaluate the integrated effects of transport on extra-tropical trace gas composition.

5.3.1 Integrated processes affecting extra-tropical composition

5.3.1.1 Mid-latitude Mean Age

Mid-latitude mean age is influenced by the ascent rate in the tropics, the strength of mixing across the subtropical barrier, and the strength of polar descent and degree of vortex isolation. In Section 5.2.1.2, mid-latitude mean ages between 10-90 hPa derived from balloon-borne CO_2 and SF_6 profiles from 35°N-45°N (Engel *et al*., 2009) were combined with tropical mean age profiles to assess tropical ascent rates. The NH mid-latitude mean age is evaluated on 4 levels in the LS (50-90 hPa) and 4 in the MS (30-10 hPa). In the SH, the only mid-latitude mean age data available are for ~50 hPa (Figure 5.5, lower right panel); these data are used in the calculation of the average mean age grade, defined in the Section 5.5.2.. The essential features of the NH mid-latitude mean age profile are a rapid increase in age from 1 to 4.5 years through the LS and nearly constant age from 24-32 km (~30-7 hPa). Below 70 hPa, most models fall within the uncertainties of the observed mean age, but above 40 hPa, 10 CCMs have ages completely outside the 1σ uncertainty range. Eight are too young and two are too old.

AMTRAC3, CNRM-ACM, NiwaSOCOL, and SOCOL have the youngest LS mid-latitude age profiles. CAM3.5, UMETRAC, and UMSLIMCAT are slightly older but still 1-1.5 years younger than the observationally-derived profiles shown in Figure 5.5. Many of these models have been diagnosed with fast tracer-derived tropical ascent and some have indications of excessive vertical diffusion, both of which likely play a role in their young mean ages here. CMAM, GEOSCCM, and MRI have the best agreement over the entire altitude range. WACCM is slightly young and ULAQ shows mixed agreement with altitude. The two UMUKCA models are 1-1.5 years older than observed, and their age is explained by slow tropical ascent.

The agreement between models and observations in the SH mid-latitudes is very similar: The same four models plus UMETRAC have the youngest ages as in the NH mid-latitude comparison. The two UMUKCA models are ~2 years older than the observations indicate. CMAM is slightly young, but the remainder of the models show good agreement. Mid-latitude mean age grades play an important role in calculating each model's globally averaged mean age grade (Section 5.5.2).

5.3.1.2 Fractional Release of Cl_y

The fractional release of a long-lived source gas such as CF_2Cl_2 (CFC-12) or $CFCl_3$ (CFC-11) is a measure of how much of the source gas has been photolysed or oxidised since entry in the stratosphere. It is defined as $f_r = (1 - X/X_{entry})$ where X is the mixing ratio at a particular location and X_{entry} is the value at the time of stratospheric entry. As CFC photolysis rates are a function of altitude and latitude, fractional release depends strongly on transport pathways in a model. Schauffler *et al*. (2003), using trace gas measurements made by instruments on the ER-2 during the NASA SOLVE campaign, show a compact relationship between the mean age of air and the fractional release for CF_2Cl_2 and $CFCl_3$. This compact relationship is robust in the lower stratosphere, and does not depend on the latitude and altitude of the measurements up to the maximum altitude of the ER-2 (~50 hPa). The mean age of a parcel is the average over all of the elements (*i.e*., transport pathways) that contribute to the parcel in question. Hall (2000) shows that in general, elements with older ages have ascended to higher altitudes, experiencing more rapid photolysis of long-lived gases than elements that have not been transported above the ozone maximum. Douglass *et al*. (2008) show that the local mixing ratio (and thus the fractional release) depends on the maximum altitude obtained by various elements of the age spectrum. The simulated relationship between fractional release and mean age thus depends on several aspects of the simulation, including the age spectrum, the maximum altitude reached by the older elements in the age spectrum, and the photolysis field.

The simulations analysed by Douglass *et al*. (2008) all used the same photolysis code, and differences in the simulated relationships between fractional release and mean age were clearly the result of differences in transport. Because the simulations analysed here did not use a common set of photolysis rates, and some use a parameterised fractional release rate based on age of air, a quantitative transport diagnostic cannot be calculated; differences in photolysis and transport both contribute to differences in the fractional release relationship. In general, simulations with greater values of fractional release for a given mean age can be thought of as having more photolytic loss, and, since the atmospheric burden of CFCs is almost completely controlled by the tropospheric burden and the boundary conditions, these simulations are associated with shorter CFC lifetimes. The evolution of active chlorine in these models would differ widely if they employed flux boundary conditions (controlling the atmospheric input) rather than mixing ratio boundary conditions (controlling the atmospheric burden).

Ten CCMs had the necessary CFC and mean age output for this diagnostic. The simulated relationship between fractional release and mean age is shown for CF_2Cl_2 and $CFCl_3$ in **Figure 5.10**. For CNRM-ACM, MRI, SOCOL, and WACCM, the simulated values of fractional release for CF_2Cl_2 are significantly greater than observed for a specified mean age. Since fractional release is strongly altitude

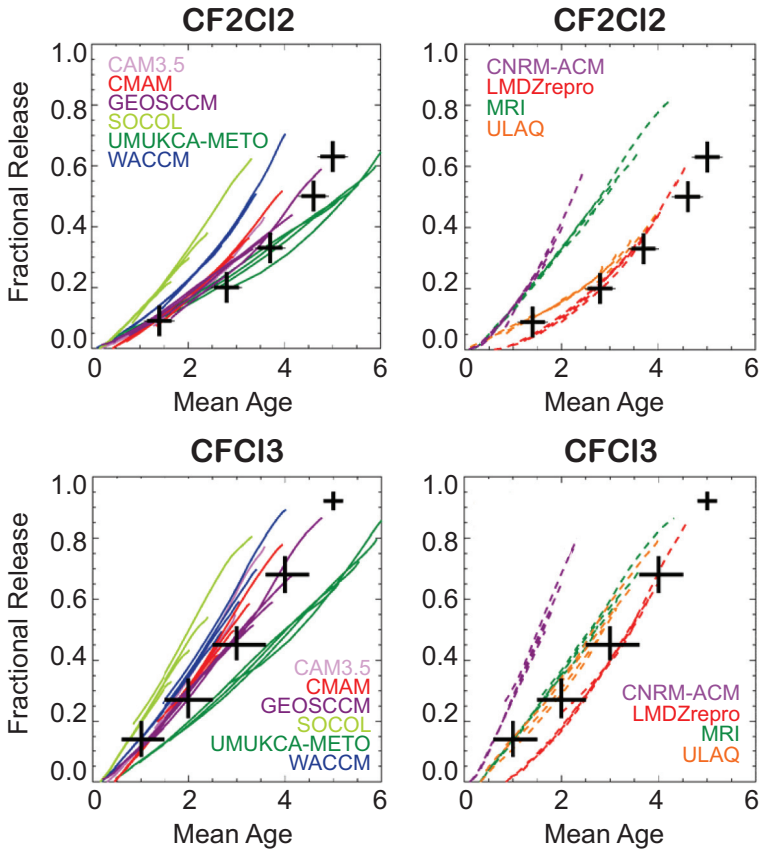

Figure 5.10: Fractional release of inorganic Cl (Cl_y) as a function of mean age of air in the lower stratosphere, where mean age was derived from CO_2 measurements. The observations used include a large suite of chlorine-containing organics and CO_2, measured simultaneously by ER-2 instruments. Schauffler et al. (2003) derived the empirical relationship between fractional release of Cl from these species and mean age and their results, including uncertainties, are plotted with large crosses. Model curves that are steeper than observed indicate that more Cl_y is released for a given mean age than observed. If all models use the same photolysis rates, this diagnostic reflects only differences in transport.

dependent, the curves falling to the left of the observations suggest that these models quickly transported air to high altitudes. For CNRM-ACM and SOCOL, this is consistent with diagnoses of fast net vertical transport in Section 5.2. However, LS ascent in MRI and WACCM was found to be fairly good, so differences in photolysis rates may play a role; see Chapter 6.3.1 for details. Interestingly, MRI is the only model to have nearly identical fractional release curves for CF_2Cl_2 and $CFCl_3$, while the empirical curves (Figure 5.10) show $CFCl_3$ to be released at much younger mean ages than CF_2Cl_2. This suggests that MRI may use the $CFCl_3$ photolysis rate for CF_2Cl_2.

Models that have young mean age at all latitudes in the lower stratosphere do not span the same range of observed mean ages (1-4.5 years); this is seen in CAM3.5, CMAM, CNRM-ACM, and SOCOL. Only UMUKCA-METO has curves falling to the right of the observations, meaning less photolysis of CFCs for a given mean age. This is consistent with the diagnosis of slow ascent, implying long transport times to high altitudes. The models that best show the observationally-derived relationship between fractional release and mean age are CMAM, GEOSCCM, LMDZrepro, and ULAQ. It is interesting that LMDZrepro agrees poorly at young ages – this is consistent with very slow ascent rates diagnosed from the age gradient in the lowest levels of the tropics. For $CFCl_3$, all of the models except MRI respond in the correct sense, *i.e.*, fractional release values for a given mean age are larger than for CF_2Cl_2, given the more rapid photolysis rate for $CFCl_3$.

5.3.1.3 Northern mid-latitude Cl_y time series

This diagnostic evaluates the time evolution of the annual mean inorganic chlorine (Cl_y) in the NH mid-latitude LS. As described in the previous section, Cl_y in the mid-latitude lower stratosphere depends on the photolysis rates of the major organic species, such as CFC-11 and CFC-12, and on the mean age of air because it is an indicator of the

maximum altitude to which a parcel has travelled. If the same photolysis rates were used in all CCMs, Cl_y would be a diagnostic that compares only transport between the models; however, this is not the case for all the CCMVal-2 models. The disagreement between the models and the observations depends on both the photolysis fields and transport differences.

The estimates of observed Cl_y are based on an analysis of UARS HALOE and Aura MLS HCl measurements (Lary *et al*., 2007). The observed mean values and uncertainties are the same as those used in WE08. The vortex grade is the average of the comparisons in October 1992 and 2005, and the mid-latitude grade is the average of 7 annual means. In the models, all chlorine comes from long-lived source gases emitted at the surface, primarily CF_2Cl_2 and $CFCl_3$, but other HCFC and CFC species contribute nearly 50% of the total Cl emitted. Maximum Cl emissions occurred in the mid-1990s and equaled roughly 3.7 ppb Cl.

Figure 5.11 shows the time series of annual mean zonal mean (35°N-60°N) Cl_y at 50 hPa for 18 models and the observations. All models show increasing Cl_y from 1960-1990, and all but UMUKCA-METO show Cl_y leveling off by the mid- or late 1990s in the lower stratosphere. (The UMUKCA-METO increase in the late 1990s and its high Cl_y mixing ratio are caused by a known error in the model's HCl washout; see Chapter 2.) The observations cover the time period from 1992 until 2004 with values between 1.7 and 1.9 ppb. There is considerable model spread, but nearly all models fall within the uncertainty of the measurements. CAM3.5 is the only model that is consistently below the uncertainty range of the observations; it has young mid-latitude mean age but a good fractional release curve, so the low Cl_y may simply reflect young age. Young mid-latitude mean age may also be to blame for low Cl_y in AMTRAC3. CMAM Cl_y is also quite low, only getting within the uncertainty range for a few years, which is surprising given its good mean age and reasonable fractional release curves. MRI and UMUCKA-UCAM are the only models with Cl_y consistently higher than observed, lying outside the uncertainty range in most years. For the

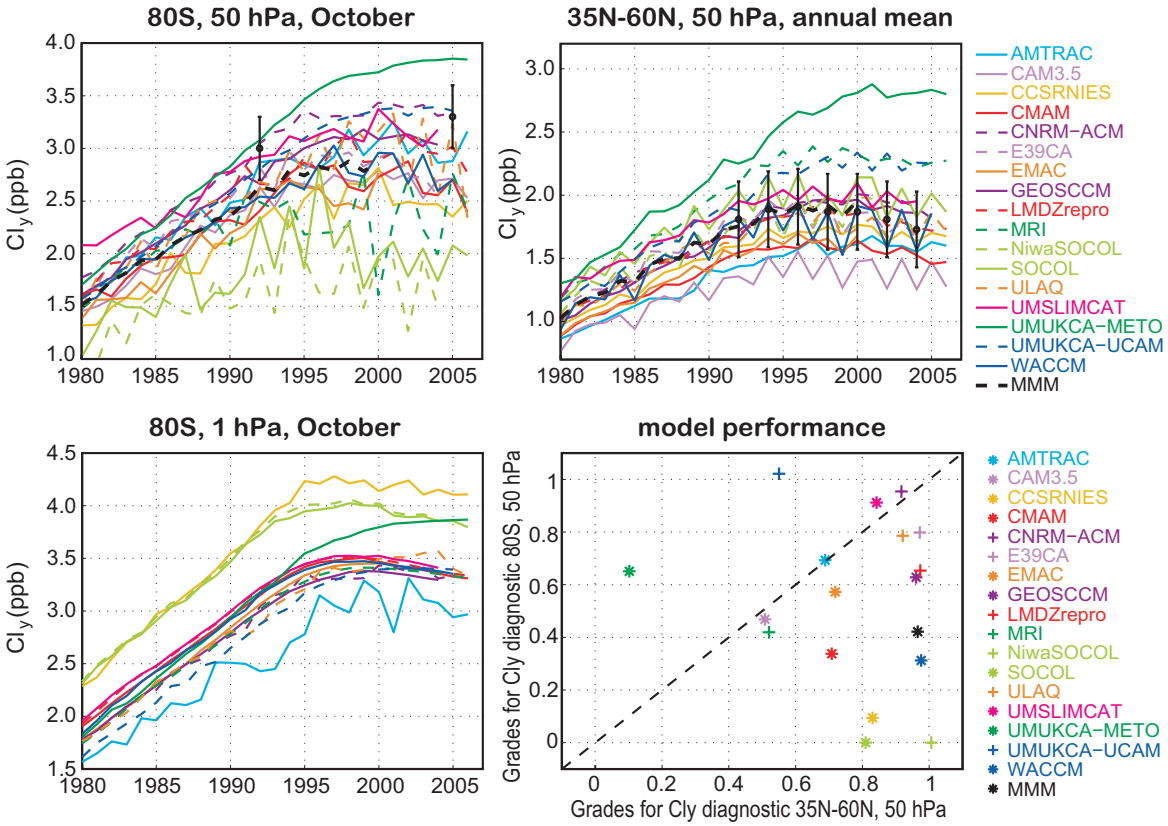

Figure 5.11: Modelled and observed Cl_y times series for 1980-2006. Estimated observational uncertainties are shown by the vertical black bars. Upper left: 80°S October Cl_y time series, 50 hPa. Upper right: NH mid-latitude Cl_y time series, 35°N-60°N, 50 hPa. Lower left: 80°S October Cl_y at 1 hPa. Lower right: grading summary for mid- and high latitude Cl_y diagnostics. Cl_y at 1 hPa is not a graded diagnostic but is included as a check for maximum Cl_y in a model. At 1 hPa, nearly all Cl-containing organics have been photolysed. Cl_y at this level should never be greater than the total Cl emitted at the surface. The prescribed REF-B1 and REF-B2 source gas boundary conditions do not exceed ~3.8 ppb in any year.

latter this may be related to very old mid-latitude mean age that has permitted air parcels more time for photolysis, but for MRI, which has good mid-latitude mean age, the high Cl_y may indicate a chemical problem. The MRI CFC-12 fractional release curve (Figure 5.10) lies well to the left of the observations, indicating much greater photochemical release of Cl for a given mean age.

This diagnostic was applied to CCMVal-1 models in WE08. CCSRNIES, LMDZrepro, and UMSLIMCAT have virtually the same mid-latitude Cl_y as they did in CCMVal-1. CMAM, SOCOL, ULAQ, and WACCM all increased a few tenths of a ppb since CCMVal-1, improving their agreement with observations. GEOSCCM also increased slightly but its agreement is the same. In CCMVal-1, MRI was too high and its Cl_y increased in the 2000's instead of levelling off. Now it levels off appropriately but it is still too high by ~0.5 ppb in the 2000's. AMTRAC3 was unrealistically high, 3 ppb, but is now slightly lower than observed (~1.6 ppb). E39CA was ~1 ppb low in CCMVal-1 but now agrees very well with observations. However, E39CA's Cl_y can not be evaluated as a transport diagnostic because the abundance is influenced by the Cl_y boundary condition imposed at the model lid (10 hPa).

5.3.1.4 N_2O annual cycle in the LS

This diagnostic assesses whether a model represents the observed balance of seasonally varying transport processes affecting the mean composition of air in the descending branch of the Brewer-Dobson circulation at 50 and 100 hPa. In the extra-tropics, increasing N_2O in spring and summer is the result of quasi-horizontal transport of young air from low latitudes, while decreasing N_2O in fall and winter shows the influence from descending, photochemically aged air. The balance of these transport processes is evaluated using monthly mean changes in the mean extra-tropical N_2O. Monthly tendencies are computed using the zonal monthly area-weighted means of Aura MLS N_2O at 50 and 100 hPa between 45°-89° in each hemisphere, averaged over a 4-year period (9/2004-8/2008). The models' outputs are normalised to a surface value of 320 ppb for consistency with the Aura MLS data set used. Because MLS N_2O has a vertical resolution of ~3 km, it was uncertain whether a model diagnostic calculated from only a single level would accurately reflect the vertical sensitivity of the MLS measurements. As an experiment, model levels above and below the MLS levels were combined in a weighted average reflecting the sensitivity of MLS retrieval to nearby pressure levels (Livesey *et al*., 2007). It was found that this averaging was unnecessary and that models' monthly trends were nearly the same as when calculated using a weighted mean on a single pressure level (50 or 100 hPa). Model behaviour is evaluated with monthly tendencies and the overall grade reported is the mean of 12 monthly grades. The standard deviations were calculated for each monthly mean from daily data and include inter-annual variability for the 4 years of MLS data used.

The N_2O monthly tendencies for 50 hPa and 100 hPa from the models and MLS with 1σ uncertainty are shown in **Figure 5.12** (SH) and **Figure 5.13** (NH). The models' performance at 50 hPa is remarkable: in both hemispheres, all models are able to realistically simulate the seasonal cycle. At 100 hPa in the NH, 15 models perform well while UMUKCA-METO, UMUKCA-UCAM, and GEOSCCM cycles indicate too much low latitude influence too early in spring. In the SH at 100 hPa, all models perform adequately in most months, although nearly all models miss the decrease in N_2O in spring and the large increase that follows in early summer (December), suggesting that low N_2O from the Antarctic vortex does not have enough impact at 100 hPa. The two UMUKCA models, EMAC, and ULAQ are the only models that produce the sharp rise in N_2O seen in early summer, but they all start to increase two months early. The UMUKCA models may capture this feature because of the very old air found in the Antarctic vortex, which gives them unrealistically low N_2O. When the vortex breaks down in late spring/early summer, the much younger low latitude air (*i.e*., with high N_2O) has a larger impact on the mean value due to the large contrast in mixing ratios.

Overall, models behave very well at 50 hPa, and although they do not perform as well at 100 hPa, all models perform roughly the same there. The seasonal cycles are worst in the SH at 100 hPa, where most models lack a low latitude (quasi-horizontal) influence in early fall, as well as strong downward influence in mid-spring. This suggests that the strength of the circulation is too weak in the lowest levels of the SH stratosphere. While this diagnostic may be useful for identifying large problems in the seasonality or strength of the Brewer-Dobson circulation, in the case of the 18 CCMs evaluated here, this diagnostic shows that seasonality of the LS circulation is quite similar in all models.

5.3.1.5 Mean age at 60°N/S

The mean age of air at 60° is influenced both by descent of older air (during fall and winter) and by horizontal mixing with mid-latitude (younger) air at a latitude that is generally outside the vortex. The observations used to derive the 60°N mean age were collected in all seasons and reflect both of these transport processes. The 60°S mean age was derived from observations between April and October, and thus has less influence from younger, low latitude air. The diagnostic is based on the mean ages and uncertainties from Andrews *et al*. (2001), and model mean ages are averaged over 55°-65° in each hemisphere. Low mean age

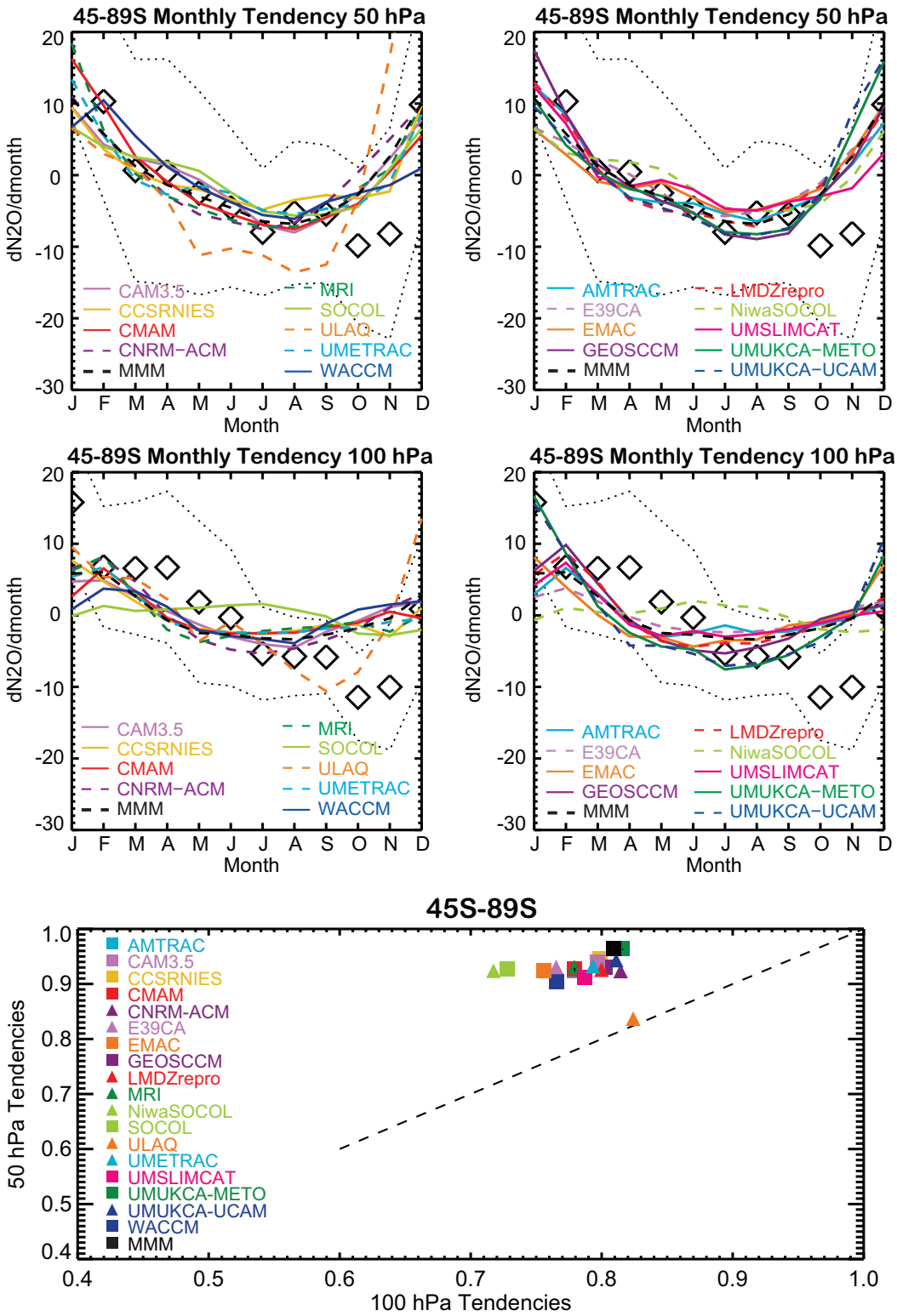

Figure 5.12: Area-weighted mean over 45°S-89°S of monthly mean N_2O tendencies at 50 hPa and 100 hPa. The monthly means and 1σ uncertainties for the observations, shown with black diamonds and dotted lines, are calculated from 4 years of Aura MLS N_2O observations. The mean annual cycle in the extra-tropics shows the seasonally varying balance between descent due to the Brewer-Dobson circulation and quasi-horizontal transport of young (high N_2O) air from the tropics. All models perform well for most of the year but show deficiencies in the austral spring.

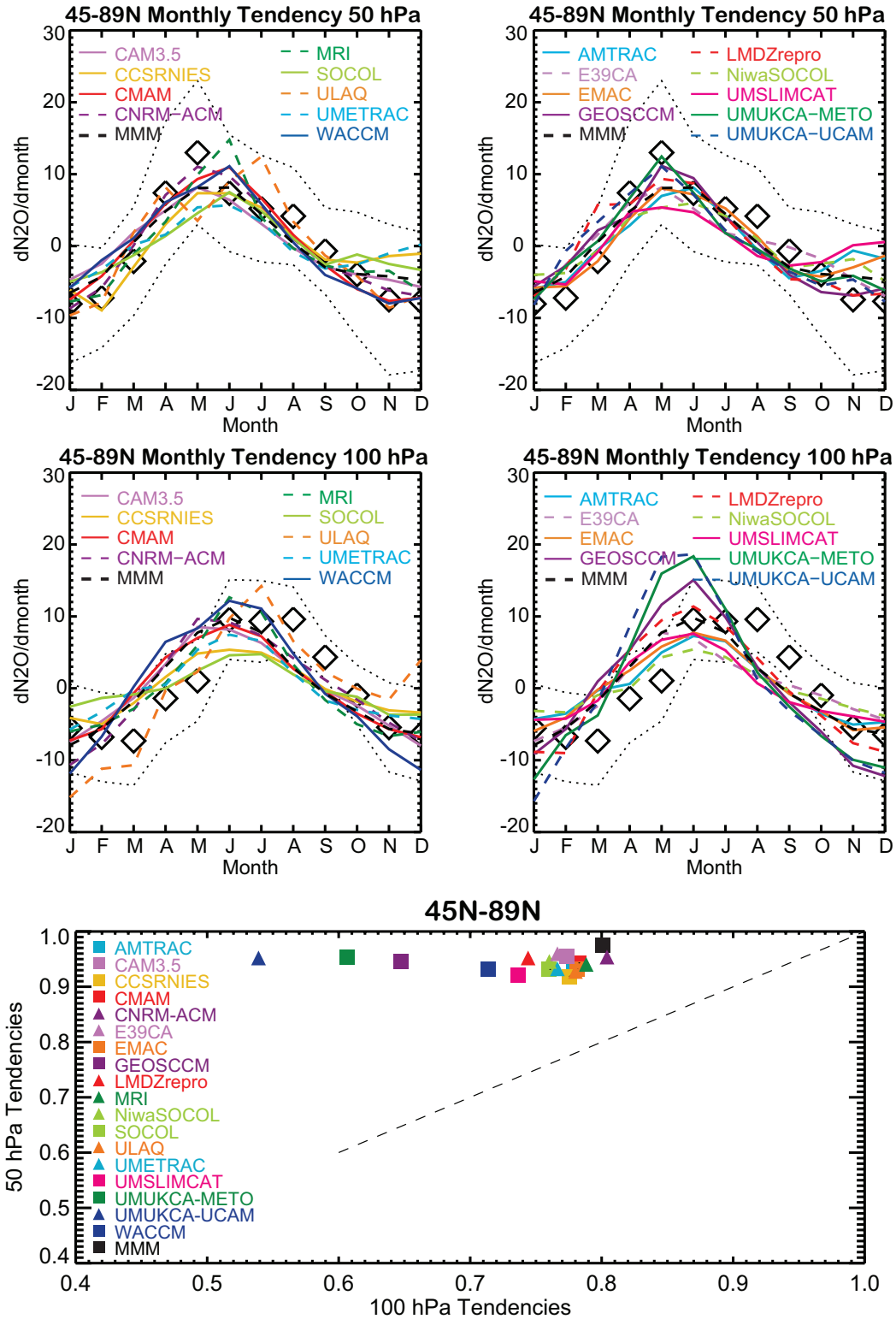

Figure 5.13: Same as Figure 5.12, but for 45°N-89°N. All models perform very well at 50 hPa in the NH, but not as well at 100 hPa. The late winter/spring N_2O increase occurs earlier than observed in several models and all show a summer time decrease that occurs two months too soon. This may impact the seasonally varying composition of the lowermost stratosphere.

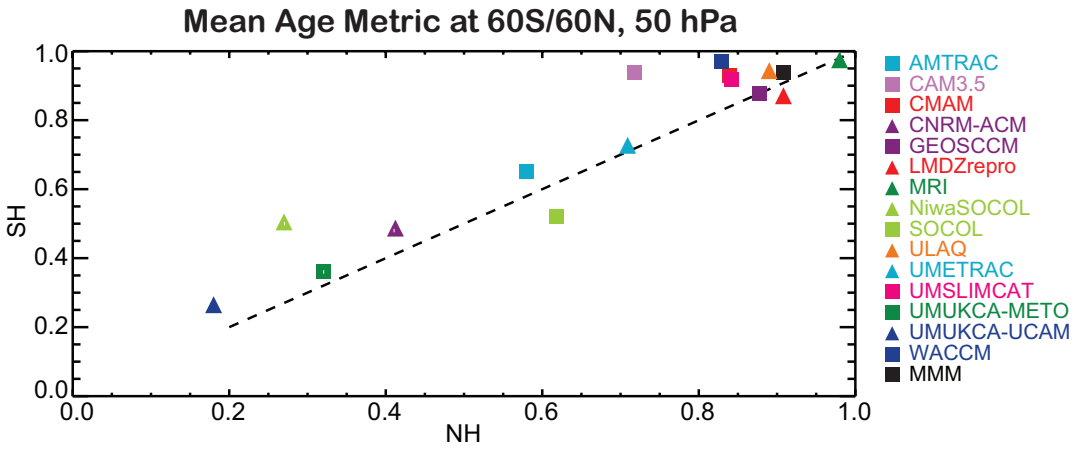

Figure 5.14: Performance metrics for model mean age of air at 60°N and 60°S, 50 hPa. About half the CCMs have mean age in close agreement with the observations. Most models perform equally well in both hemispheres.

at high latitudes may be an indication of weak descent, poor vortex isolation, or fast circulation, all of which can lower mean age everywhere. The models' results, shown in **Figure 5.14**, illustrate that the models perform equally well in both hemispheres.

Model and empirically derived mean age as a function of latitude at 50 hPa is shown in Figure 5.5. CNRM-ACM has the youngest age, 3 years or less at all latitudes. NiwaSOCOL, SOCOL, and AMTRAC3 are slightly older but are 1-1.5 years younger than observed at 60°. All four have polar ages outside the uncertainties and all except AMTRAC3 have indications of excessive vertical diffusion. The two UMUKCA models have been diagnosed with slow circulations and their mean ages are much greater than observed in both polar regions (> 6 years). The nine remaining models that can be evaluated fall within the uncertainty of the observations most of the time. The best mean ages are found in CAM3.5, CMAM, GEOSCCM, LMDZrepro, MRI, ULAQ, UMSLIMCAT, and WACCM.

Mean age at 60°N/S was not explicitly evaluated in CCMVal-1, however, comparison to Figure 10 in Eyring *et al*. (2006) shows that the bulk of the models have 60°N/S mean ages very similar to the age shown in this evaluation (Figure 5.5). The models with the most noticeable change at these latitudes are UMETRAC (was ~3 years, now has increased by ~0.5 year), MRI (decreased from ~6 years to ~4.5 years), and UMSLIMCAT (was > 5 years at 60°S but it now just under 4 years). These changes improved the agreement with observations for UMETRAC and MRI.

5.3.2 Polar processes

5.3.2.1 Antarctic Spring CH_4 PDFs

This diagnostic uses 9 years of Antarctic UARS HALOE CH_4 and temperature profiles to examine the degree of vortex isolation in early spring at 13 levels from 400 to 2000K (~68-1 hPa); it is similar to the vortex isolation diagnostic developed in Strahan and Polansky (2006). The HALOE data set used spans the period 1993 to 2001 and includes profiles from 68°S-78°S from October to mid-November. Because of HALOE's small daily spatial coverage it is not possible to calculate an area-weight probability distribution function (PDF). The HALOE data shown are simply distributions of the observed mixing ratios within the noted latitude range, which is not uniformly sampled. The months chosen for this evaluation are constrained by the sampling pattern of the HALOE instrument and the UARS orbit; latitudes poleward of 60° are sampled only in two seasons per year in each hemisphere. The models are evaluated over a relatively large latitude range (50°-80°), so that a model having a particularly large vortex will not get a low score simply because the location of the barrier is equator-ward of 68°S. All models except EMAC show a transition to easterlies at 60°S occurring after October (see Figure 4.2), so this evaluation is made before vortex breakdown occurs. For EMAC, the October transition to easterlies occurs only in the upper stratosphere.

All HALOE high latitude CH_4 data from the 9-year period were interpolated to 13 isentropic surfaces and then binned to create the contoured distributions shown in **Figure 5.15**. Where the distributions are bimodal, the peak with the lower mixing ratio indicates vortex air, and the other peak identifies the most probable mid-latitude mixing ratio in this latitude range. PDFs for the models are calculated at each of 13 isentropic levels and are defined as bimodal if both peaks have a Gaussian shape. In Figure 5.15, the key features shown by the HALOE vortex distribution at 700 K and above are a lack of a vertical gradient, a narrow distribution, and low mixing ratios, all of which indicate very isolated descent and little interannual

variability. Below 700 K, bimodal distributions have deep minima, indicating that the vortex is isolated, but vortex air found at these levels has greater influence from younger air (*i.e.*, higher CH_4).

The HALOE most probable value (MPV) profiles and those from 14 CCMs are shown in **Figure 5.16**. Grading is based on the difference between mid-latitude and vortex MPVs (MPV2-MPV1, or ΔMPV). When a model distribution has a single peak, ΔMPV is set to zero. Because interannual variability in the vortex is low, and the quantity evaluated is the difference between two measurement means, the HALOE ΔMPV uncertainty is quite low, probably less than 10%. The diagnostic identifies whether the mid-spring Antarctic vortex has an appropriate barrier to transport at 13 levels from 450-1300 K, based on the separation of the mid-latitude and vortex MPV profiles. At 400 K and above 1300 K, the distribution has only 1 peak and ΔMPV is zero. At 400 K the single peak distribution is

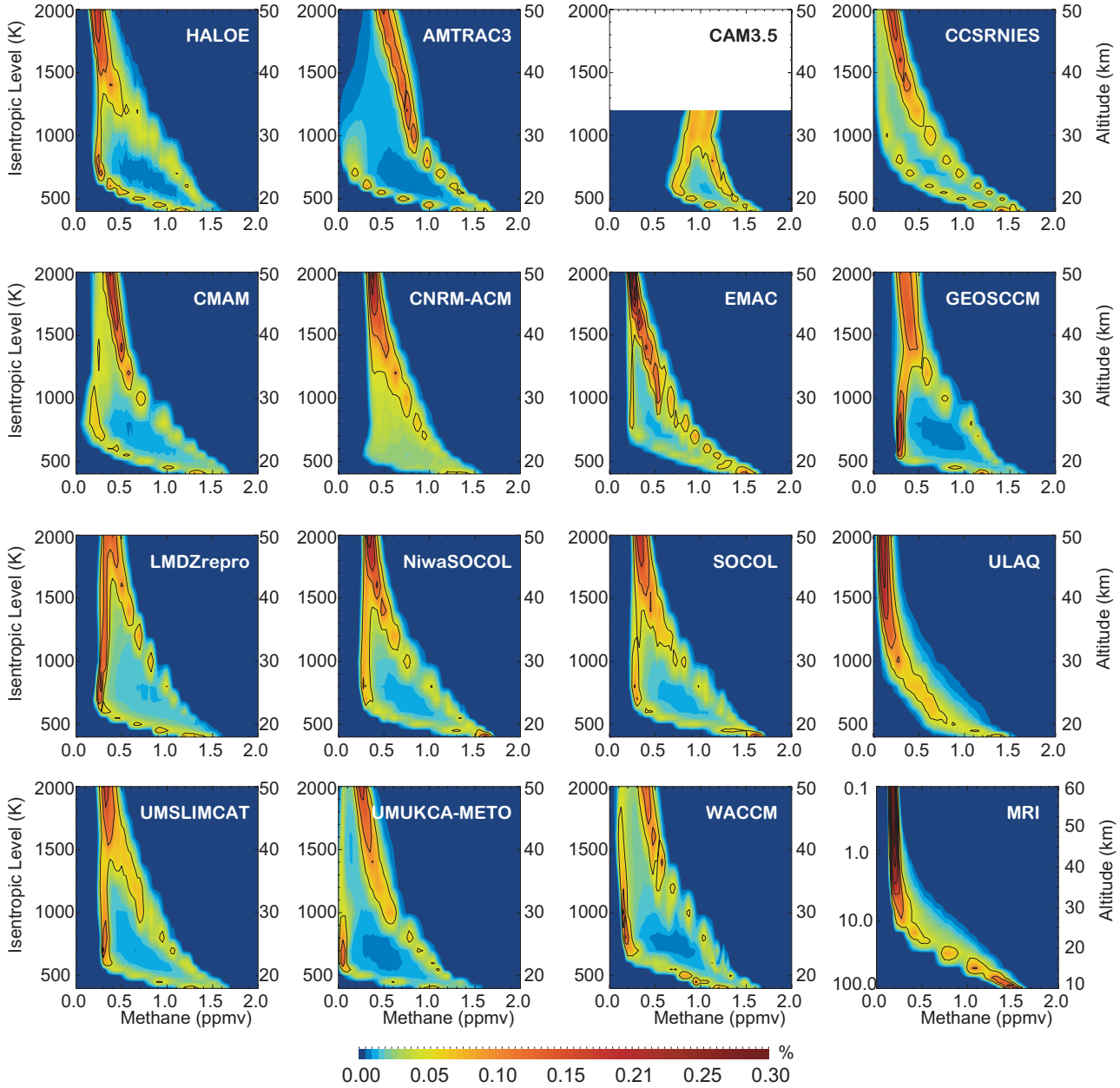

Figure 5.15: Contoured PDF of HALOE CH_4 data, 68°S-78°S, for the period mid-October to mid-November (first panel). Model PDFs are calculated from 50°S-80°S (see text). Yellow and red indicate the most probable values, blue is the least probable value. The HALOE distributions are bimodal from 450-1300 K, indicating low-CH_4 vortex air that is isolated from the mid-latitudes (the high CH_4 branch of the PDF). MRI model results (bottom right panel) are shown on pressure surfaces.

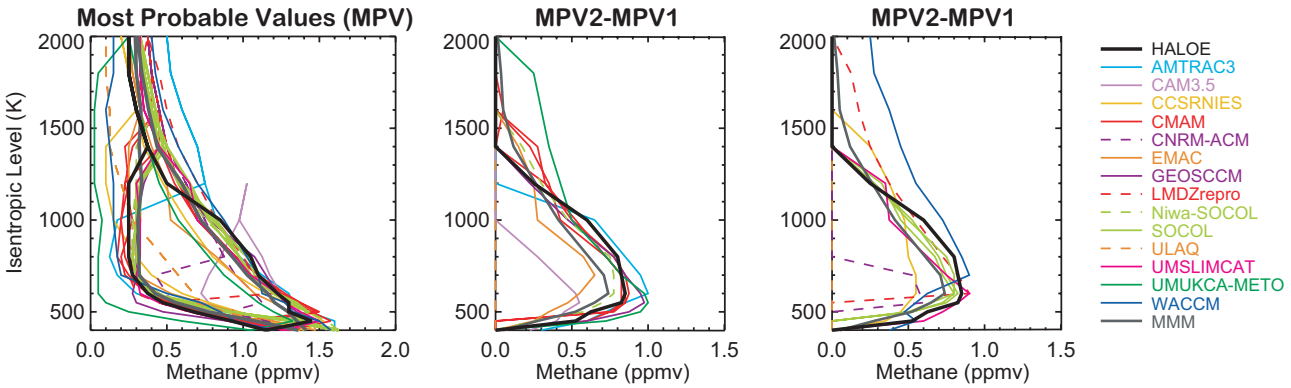

Figure 5.16: The most probable values (MPVs) of the CH_4 PDFs identified from HALOE and model analyses (left panel). The grading metric for this diagnostic is based on a model's agreement with difference between the MPVs in the mid-latitudes and vortex (MPV2-MPV1, or ΔMPV). The centre and right panels show the ΔMPV profile from HALOE, 14 CCMs (split into two panels), and the multi-model mean (MMM).

characteristic of vortex air, and at levels 1400 K and above, ΔMPV=0 implies that spring planetary wave activity has eliminated the transport barrier, mixing vortex and mid-latitude air together. Vortex isolation is particularly important at levels where large spring-time O_3 loss occurs, so grading is divided into a LS grade for PSC-forming levels (400-700 K) and a middle and upper stratospheric grade above that (800-2000 K).

The CH_4 PDFs from 15 CCMs are shown in Figure 5.15. In the LS, CNRM-ACM, MRI, ULAQ show no evidence of a transport barrier. (MRI could only be evaluated on pressure surfaces, but its PDFs show single mode behaviour at all levels. It is highly unlikely the evaluation would be any different on isentropic surfaces.) CAM3.5, CCSRNIES, and EMAC show an LS barrier but with too little separation between MPVs. NiwaSOCOL and SOCOL have reasonable LS barriers at 500 K and above, but do not extend to 450 K. Poor agreement in the 450-550 K range can have a serious impact on a model's ability to sequester high Cl_y inside the vortex. LMDZrepro has some LS barrier but the predominance of low CH_4 suggests the vortex is extremely large. A good LS barrier is found in AMTRAC3, CMAM, GEOSCCM, UMSLIMCAT, UMUKCA-METO, and WACCM.

In upper stratosphere around 1400 K, the observations show a shift from bimodal to single mode PDF. Some models do not maintain a vortex barrier to high enough levels (AMTRAC3 and CAM3.5) and some have a vortex that persists to very high levels in the upper stratosphere or lower mesosphere (CCSRNIES, LMDZrepro, UMUKCA-METO, and WACCM). The AMTRAC3 PDF is single mode above 1000 K, but the mixing ratios are much higher than observed, suggesting that the vortex above 1000 K broke down and became well-mixed with mid-latitude air by October. Although UMUKCA-UCAM did not submit the required output for this test, its circulation is very similar to UMUCKA-METO and probably has very similar vortex representation.

Because of large interannual variability in the Arctic vortex size and duration, the HALOE high northern latitude observations, primarily from March and April, do not regularly sample the vortex when it is well-isolated. These data cannot be used to derive a robust diagnostic of Arctic vortex isolation in late winter.

5.3.2.2 Antarctic September N_2O profiles

This diagnostic assesses the integrated effects of descent and mixing deep in the vortex during the late austral winter. It uses 5 years of Aura MLS v2.2 N_2O September observations from 80°S-88°S between 10-100 hPa (Livesey *et al*., 2007). From 10-68 hPa, the N_2O accuracy is 9-14%, but 25% at 100 hPa. September was chosen for the analysis because descent that has been occurring since autumn is essentially complete, but in the lower stratosphere the vortex is still very strong and un-influenced by recent transport from lower latitudes. The observational uncertainties used are calculated from measurement accuracy and the interannual variability in the N_2O data. The diagnostic is calculated on 7 pressure levels between 100 and 10 hPa; the reported grade is the average of grades from these levels.

Figure 5.17 shows the observed and model N_2O September mean profiles in the Antarctic. Most of the CCMs fall within the one sigma uncertainty for many or all parts of the profile. Seven CCMs lie significantly further from the observations. CAM3.5, CCSRNIES, E39CA, NiwaSOCOL, and SOCOL have N_2O mixing ratios much higher than observed, suggesting either insufficient vortex descent and/or insufficient isolation from mid-latitudes. In the cases of E39CA and CAM3.5, the high N_2O is likely a consequence of their low lids (10 and 3 hPa, respectively). Insufficient isolation has been diagnosed for CAM3.5 and CCSRNIES. NiwaSOCOL and SOCOL show very good isolation except at 450 K, so their high N_2O may be the

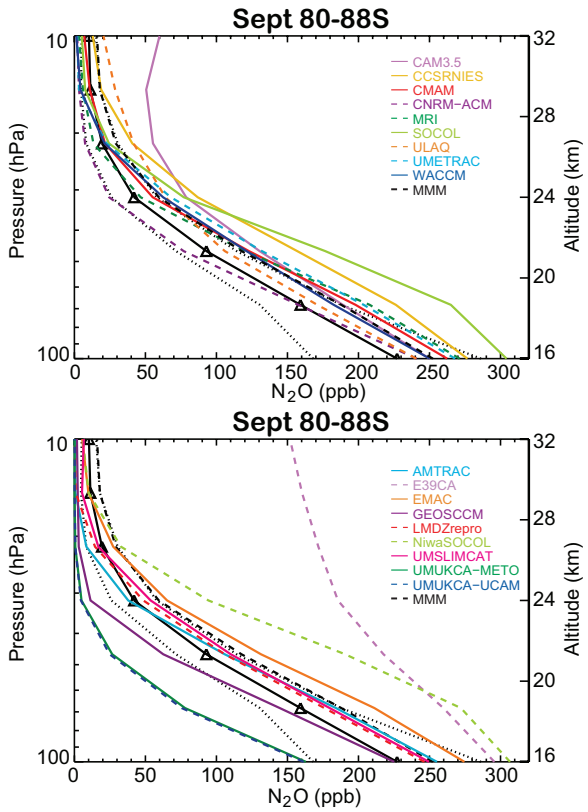

Figure 5.17: 18 CCM and observed profiles of N_2O, 80°S-88°S, for September. The observed mean profile (triangles) was derived from 5 years of September Aura MLS observations; the combined effects of measurement uncertainty and interannual variability (1σ) are shown by the black dotted lines. The September N_2O profile at these latitudes reflects the net effect of descent and mixing that occurred in the vortex during winter.

related to their overall low mean ages and may again reflect excessive vertical diffusion. Both of the UMUKCA models have N_2O much lower than observed, consistent with very old mean age in the vortex.

5.3.2.3 Antarctic spring Cl_y time series

The time evolution of Cl_y for the Antarctic in October was evaluated in WE08. The estimates of observed Cl_y in the Antarctic stratosphere are based on HCl measurements from HALOE in 1992 (Douglass *et al*., 1995; Santee *et al*., 1996) and from MLS on the Aura satellite in 2005 (WMO, 2007); the observations show evidence for the complete conversion of Cl_y (HCl and $ClONO_2$) to ClO_x in the deep vortex by early October. By the end of October, ClO_x has converted back to HCl, not $ClONO_2$, because the vortex is denitrified. Thus, the observed HCl mixing ratio in late October can be used as an estimate of vortex Cl_y (Douglass *et al*., 1995). The uncertainties used for the Cl_y estimates are discussed in Eyring *et al*. (2006). Because the dates and latitudes sampled by HALOE change every year, the estimation of Cl_y from HCl was only possible in 1992. In all other years deep vortex air was not sampled during the brief period when all ClO_x has been converted to HCl. October vortex Cl_y was estimated from Aura MLS HCl data for 2005 only because of a radiometer failure in 2006.

The upper left panel of Figure 5.11 shows the evaluation of October monthly mean model Cl_y at 80°S in 1992 and 2005. The results reported represent the average of the grades for both years, or only for the 1992 grade if the simulation did not continue to 2005. Low Cl_y in the vortex can result from insufficient vortex isolation that allows low Cl_y from the mid-latitudes to mix into vortex, or from a circulation that does not transport source gases to high enough altitudes for the CFC photolysis necessary for Cl_y production.

Except the two UMUKCA models, all models show a Cl_y increase beginning in the 1980s that levels off by the late 1990s. UMUKCA-UCAM flattens in the early 2000's, consistent with its mean age being several years older than the other models, and UMUKCA-METO appears to peak in 2005. The unusually high Cl_y seen in UMUKCA-METO (~3.9 ppb in 2006) is a result of a known bug involving the treatment of HCl washout. MRI, NiwaSOCOL, and SOCOL show extremely large interannual variability and fall well below the observations. In the case of NiwaSOCOL and SOCOL, both show good vortex isolation except at 450 K, but have low mean ages due to rapid net vertical transport, which could explain their low Cl_y. MRI has very good mean age, but its lack of a vortex barrier (Figure 5.16) could explain the low Cl_y. Most of the models under-estimate the observations, but CNRM-ACM, GEOSCCM, ULAQ, UMSLIMCAT, and UMUKCA-UCAM fall within the uncertainty at both data points. The very good agreement of CNRM-ACM with observations is quite surprising given its poor Antarctic vortex isolation (Figure 5.15) and very young age of air at high latitudes (Figure 5.5). However, this CNRM-ACM result is consistent with their fractional release behaviour (Figure 5.10), which shows unusually large fractional release for a given mean age. E39CA's Cl_y is slightly below the observations, but as discussed in Section 5.3.1.3, E39CA Cl_y cannot be evaluated as a transport diagnostic.

Nearly all organic Cl from source gases has been converted to Cl_y by the time it reaches the upper stratosphere, thus the Cl_y time series at 1 hPa should be less than or equal to the surface organic Cl time series lagged by several years. In the REF-B1 scenario the maximum surface Cl is less than 3.8 ppb. The bottom left panel of Figure 5.11 shows model Cl_y time series at 1 hPa, 80°S. CCSRNIES, NiwaSOCOL, SOCOL, and UMUKCA-METO all have maximum Cl_y greater than the surface boundary condition. The UMUKCA problem is reported to be caused by a

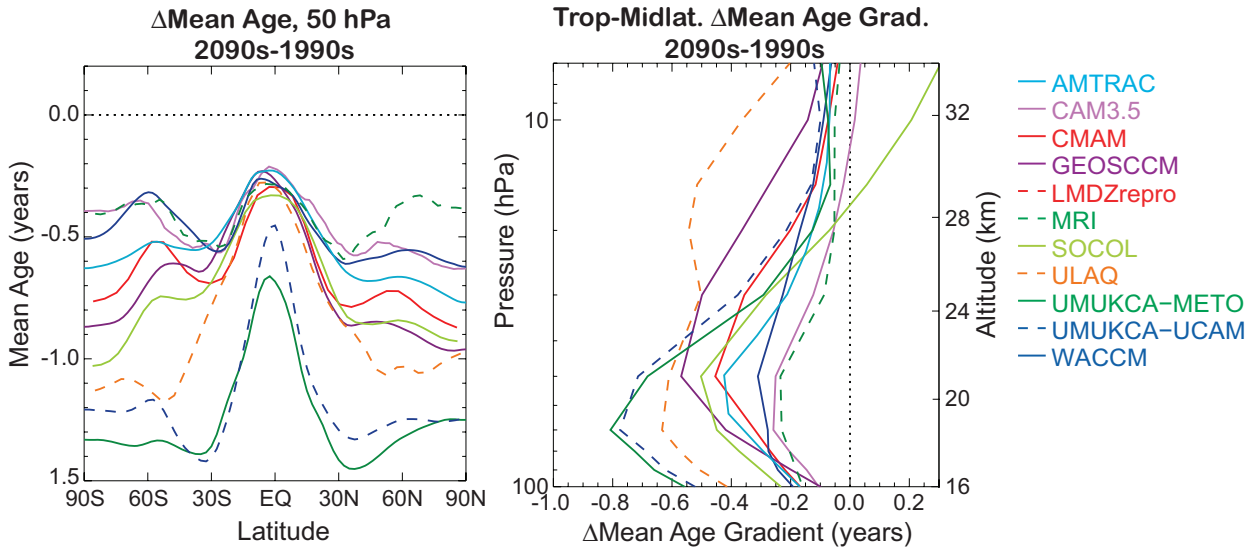

Figure 5.18: Mean age changes in the REF-B2 simulations during the 21st century. The mean age difference is the difference between an average of the last 10 years of the REF-B2 run, usually 2090-2099, and an average over 1990-1999. Ten CCMs submitted mean age for the REF-B2 scenario. The left panel shows the change in mean age at 50 hPa for the CCMs, all of which predict younger age at all latitudes at the end of the 21st century. All ten CCMs also predict that tropical ascent rates based on the age gradient will increase in the lower stratosphere (right panel), and most models predict a slight increase in age gradient ascent rates in the middle stratosphere as well.

known HCl washout problem. If the source gas boundary conditions for Cl are correctly implemented in the other three models, these results imply a lack of local mass conservation for the Cl_y family.

Antarctic Cl_y was evaluated for 13 CCMs for CCMVal-1 using the same data sets shown here and was reported in WE08. Six models have essentially the same behaviour as seen in CCMVal-1: CCSRNIES, CMAM, EMAC, LMDZrepro, MRI, and UMSLIMCAT. AMTRAC3 was significantly too high in mid-latitudes and in the Antarctic but is now much lower and agrees much better with the mid-latitude observations. Five models, GEOSCCM, E39CA, SOCOL, ULAQ, and WACCM, all have higher Cl_y and improved comparisons with observations. E39CA prescribes an upper boundary condition for Cl_x at 10 hPa, so this diagnostic is not a true transport test for this model. For SOCOL and WACCM, the Cl_y improvement is most confined to the mid-latitudes; their polar Cl_y is still low. The UMETRAC REF-B1 simulation was not available for this analysis.

5.4 Stratospheric transport changes in the 21st Century

Several diagnostics that reveal fundamental characteristics of stratospheric transport have been evaluated for the REF-B2 scenario. In this section, model simulation of the recent past, 1990-2006, is contrasted with simulation of the late 21st century, roughly 2080-2099. The diagnostics evaluated are mean age in the lower stratosphere, the tropical-mid-latitude mean age gradient, the water vapour tape recorder, and Antarctic vortex isolation.

The mean age of the lower stratosphere shows a striking and consistent change among the 10 CCMs that ran the REF-B2 scenario. **Figure 5.18** shows the difference in mean age at 50 hPa between the future and the recent past. All 10 models predict a decrease in mean age. The models also show similar latitudinal structure. All show the smallest age decrease in the tropics, and most show the second smallest decrease near 60°N/S and the greatest decreases at 30°N/S. Except for the two UMUKCA models that have by far the slowest circulation, all models make nearly identical predictions between 15°S-15°N. In the extra-tropics, the bulk of the models predict age decreases of 0.5-1.0 year.

The decrease in mean age in the lower stratosphere is consistent with faster ascent rates predicted for the end of the century, indicated by a decrease in the tropical-mid-latitude mean age gradient (right panel of Figure 5.18). All models predict an increase in the speed of the circulation in the LS, extending as low as 100 hPa, with maximum increases in ascent between 50-30 hPa. Ascent rates at 10 hPa show only a slight increase in the future. The mean age and age gradient results are consistent with Chapter 4 results showing a strong consensus among models that there will be a stronger circulation in the 21st century and with results from analyses of the future changes in the circulation and age of air from the CCMVal-1 simulations

(Li *et al*., 2007; McLandress and Shepherd, 2008; Garcia and Randel, 2008; Oman *et al*., 2009). However, the CCMVal-1 models that provided age output (AMTRAC, CMAM, GEOSCCM, and WACCM) also all simulated a decrease in the mean age of air between 1960 and the present (Waugh, 2009). These results are contradicted by the Engel *et al*. (2009) data, which show little or no trend in the age in the NH above 24 km. Nevertheless, there is indirect evidence of an increase in the tropical upwelling in the past few decades from trends in lower stratospheric ozone, water vapour, and temperature (Randel *et al*., 2006). The differences between observed and modelled age trends could result from errors in simulating compensating circulation changes (such as an increase in poleward transport in the lowermost stratosphere) or changes in mixing (which could offset a decrease in age from the circulation) and will continue to be an important area of research.

Of the 10 models that submitted REF-B2 water vapour, only six had water vapour anomalies with a clearly identifiable and easily analysed tape recorder signal in the TLS (CAM3.5, CCSRNIES, CMAM, LMDZrepro, SOCOL, and WACCM). Of these, only three had sufficient amplitude to be analysed in the TMS as well (CMAM, SOCOL, and WACCM). The base of the tape recorder is higher at the end of the 21st century than during present day in all but one model. The changes in the tape recorder phase speed (ascent rate) provide a less clear picture of changes in the circulation than the age gradient, but nevertheless all models but one (SOCOL) predict an increase in the phase speed over most of the tape recorder altitude range. It is interesting that SOCOL predicts a decrease in the phase speed when it has one of the fastest phase speeds in the present-day simulation and has indications of excessive vertical diffusion. The models also generally predict less rapid attenuation of the tape recorder signal in the future, which is consistent with the increase in phase speed and the decrease in mean age gradient.

Antarctic vortex isolation at the end of the 21st century has been examined for the five CCMs that submitted the output necessary to calculate Antarctic CH_4 PDFs. These models are CCSRNIES, CMAM, GEOSCCM, UMUKCA-METO, and WACCM. CCSRNIES, GEOSCCM, and UMUKCA-METO show increased mixing between the vortex and mid-latitudes above 1000 K in the future, suggesting that winter planetary wave activity may have increased, stirring the vortex and mid-latitudes together earlier in the season. Chapter 4 notes that SH final warmings occur earlier in the 2000-2049 period, though there was no consensus among models on the type of tropospheric wave forcing responsible. WACCM shows less vortex and more mid-latitude air in the PDFs above 1500 K, which indicates some increased mixing at high altitudes. CMAM showed a similar shift from vortex to mid-latitude values from 1000-1500 K in only 1 of its 3 ensemble members. The only thing that all of the models had in common in the future was a slightly deeper minimum separating the vortex and mid-latitude peaks in the MS (~600-1000 K). The predicted change in LS vortex size, as judged by the area under the vortex peak in the PDF, varied between the models. In CMAM and GEOSCCM, the LS vortex is predicted to increase in size, while UMUKCA-METO and WACCM each have a smaller vortex in the future. CCSRNIES did not have a clear change in size. These 5 CCMs provide no consensus as to how the Antarctic vortex size and depth may change in the future.

5.5 Summary

5.5.1 Transport Summaries by Model

Multi-Model Mean (MMM). The multi-model mean tape recorder signal was not quantitatively assessed, but the vertical velocity inferred from the mean tape recorder phase speed, which includes only the eight models for which the tape recorder could be analysed in both the TLS and TMS (and does not include E39CA at the lowest level, where its phase speed is infinite), is considerably faster than the observed tracer-derived ascent at all levels. However, it is strongly influenced by two models with very fast tape recorder phase speeds (NiwaSOCOL and SOCOL). The age gradient vertical velocity was calculated from 15 models and includes two models with very slow ascent (UMUCKA-METO and UMUCKA-UCAM), which were not available for the tape recorder analysis and which balance out the models with very fast ascent. The age gradient vertical velocity is much closer to the observations and suggests no overall bias in the CCMVal-2 models' ascent rates. The tropical CH_4 gradients indicate slightly too much mixing below ~25 hPa and slightly too little mixing in the SH winter above. Mean age at 50 hPa matches the observations closely at all latitudes. The tropical mean age profile is within the uncertainty of the observations at all altitudes while the mid-latitude mean age profile is slightly too young in the MS. These comparisons suggest, at most, only small biases in tropical ascent and tropical-mid-latitude mixing in the CCMVal-2 models as a whole.

It is interesting that the mean of all models' average mean age (AMA) grade and the AMA grade of the MMM are quite different. When averaging the mean age of all CCMs together, young mean ages compensate for old mean age to produce a multi-model mean age that scores a very high AMA grade (0.85). However, the average of individual model AMA grades is only 0.57. This quantity is a more meaningful indication of the overall transport credibility in CCMs than the grade for the physically unmean-

ingful MMM. It indicates that the average transport fidelity of the 15 CCMs is only fair.

In the Antarctic, three independent diagnostics point to insufficient vortex isolation in the mean model. The Antarctic N_2O profile is about 1σ higher than the observations, Cl_y inside the vortex in October about 1σ lower than the observations, and the lower stratospheric separation between mid- and high latitude most probable CH_4 values in October is less than observed. Overall, while the tropical and mid-latitude transport diagnostics do not indicate any obvious biases, the average of all model AMA grades suggests that transport deficiencies exist in most models. The mean model is unable to produce a sufficiently isolated, chemically perturbed lower stratospheric vortex in austral spring.

AMTRAC3: The tape recorder and age gradient vertical velocities are too fast in the TLS, but close to observed in the TMS. The tape recorder attenuation and CH_4 gradients indicate somewhat rapid mixing in the TLS, but there are significant tracer-dependent differences in mixing diagnostics in the TMS, with the tape recorder attenuation indicating somewhat too much mixing while the CH_4 profiles have very weak vertical gradients, indicating too little in-mixing from mid-latitudes. The tropical age is very young and the shape of the age profile also suggests too little mixing in the TMS. The N_2O PDFs generally agree well with the observations in the TMS, with somewhat stronger than observed separation above 800 K. The 60°N/S mean ages are about a year too young. Antarctic September N_2O agrees well from 100-20 hPa but is too old above that, and the vortex isolation looks good in the LS up to 1000 K. No fractional release could be calculated. The Cl_y time series in the mid-latitudes is lower than observed, consistent with mean age that is more than 1 year too young. The Antarctic 50 hPa Cl_y is only slightly low but it is higher than its value at 1 hPa in some years. This is consistent its mean age in the Antarctic LS being older than in the US. Furthermore, vortex PDFs above 1000 K (~8 hPa) show strong mixing between mid- and high latitudes, consistent with a reversed vertical age gradient and lower Cl_y at 1 hPa. Overall, the model is characterized by a somewhat fast circulation in the LS and good circulation in the MS. Tropical-mid-latitude (T-M) mixing is a little strong in the TLS and too weak in the TMS, although the TMS diagnostics are mixed. The AMA grade is low. There is reasonable LS vortex isolation, but the lack of a vortex above 1000 K and the inverted Cl_y and mean age vertical gradients hint at strong high altitude wave activity in the late Austral winter.

CAM3.5: The tape recorder signal could only be analysed in the TLS. The tape recorder shows somewhat rapid ascent in the TLS, consistent with the age gradient vertical velocities. The age gradient ascent is also fast in the TMS, with increasing divergence from the observations and from the residual vertical velocity with height. Most T-M mixing diagnostics support too much mixing in the TLS and TMS. The 60°N/S ages are fairly good, but Antarctic N_2O suggests there is slightly young air in the LS vortex. CH_4 and N_2O are too high in the vortex just above the LS, suggesting a problem with getting sufficient descent with isolation. The model does not show a distinct vortex above 900 K – the branches of the PDFs are merged. Mid-latitude Cl_y is considerably lower than the observations and is the lowest of the CCMs. Vortex Cl_y is also too low. The fractional release curves are in good agreement with observations. Overall, the tracer-derived circulation is somewhat fast in the LS, and more so in the MS, and T-M mixing is too strong at all levels. This model has reasonable mean age in the LS but the fast circulation gives ages that are too young in the MS in spite of the strong T-M mixing. The AMA is better than the average CCMVal-2 model. The fast vertical transport may be responsible for the low Cl_y in the mid-latitudes and vortex. It is likely that deficiencies in the circulation and mixing can be at least partially attributed to the model's low lid (~40 km).

CCSRNIES: The mean age could not be evaluated for this model. The tape recorder signal could only be analysed in the TLS. The tape recorder phase lag is zero above the tropopause and then agrees quite well with observations, so the overall ascent rate in the TLS is somewhat fast. The tape recorder attenuation indicates fairly rapid mixing in the TLS, and this is supported by the CH_4 profiles that fall off quite rapidly. The CH_4 gradients and N_2O PDFs also indicate too much T-M mixing in the MS. The Antarctic N_2O profile, which is too high at all levels from 100-10 hPa, is consistent with fast ascent, but this cannot be confirmed because there are no TMS ascent diagnostics. The Antarctic vortex shows some barrier to mixing between 500-1500 K, but high values of CH_4 inside the vortex (and low values outside) suggest a weak barrier. The mid-latitude Cl_y time series agrees well with observations, but in the vortex, Cl_y is too low by about 1 ppb. CFC-11 and CFC-12 J-values are well outside expected values (see Chapter 6). Cl_y at 1 hPa is greater than 4 ppb after 1994, which is higher than the total Cl prescribed by the boundary conditions in any year of this simulation. Overall this model has a slightly fast LS circulation with too much T-M mixing in the LS and MS. Antarctic composition is consistent a fast circulation. Cl_y family mixing ratios are not conserved.

CMAM: In the TLS the tape recorder phase speed is variable but overall agrees well with observations. The age gradient vertical velocity is slightly fast, but within the range of observational uncertainty. Relatively good consistency was found between the three estimates of vertical velocity in the TLS. In the TMS, the age gradient ascent rate agrees

quite well with the residual vertical velocity and with observations, but the tape recorder ascent rate is somewhat fast. The tape recorder attenuation indicates very good mixing rates in the TLS, corroborated by good CH_4 gradients and tropical mean ages. In the TMS there are significant tracer-dependent differences in the mixing diagnostics, with the tape recorder amplitude indicating far too little mixing, despite the fact that the CH_4 gradients are very close to observed and the tropical mean ages are only slightly young. The N_2O PDFs agree very well with the observations, also indicating good tropical-extra-tropical mixing. The 60°N/S mean ages are good. The Antarctic N_2O profile is a little high in the LS but agrees with observations at and above 30 hPa. The vortex isolation diagnostic shows a very good LS vortex barrier. Both descent and isolation in the Antarctic LS appear reasonable. Fractional release curves are a little higher than the observations, meaning that for a given mean age this model produces more Cl_y than observed. In spite of this, Cl_y in the mid-latitudes and vortex are both ~1σ lower than the observations. Overall, this model has reasonable ascent and good T-M mixing in the LS. T-M mixing in the MS looks very good using three diagnostics while the tape recorder attenuation shows no mixing at all. Consistent with good circulation and mixing, the AMA is better than most. The reasons for low Cl_y are undiagnosed.

CNRM-ACM: The tape recorder signal could only be analysed in the TLS. The tape recorder and the age gradient both indicate very rapid ascent in the TLS. The age gradient also indicates very rapid ascent in the TMS, and the differences between the tracer-derived velocities and the residual vertical velocity are very large, suggesting, along with other diagnostics, that there may be excessive vertical diffusion. The mean ages are very young everywhere, consistent with very fast net vertical transport. The tape recorder is very rapidly attenuated in the TLS, and the CH_4 gradients agree reasonably well with observations despite the fast tracer ascent rates, suggesting either too much mixing or vertical diffusion or a combination of both. However, the N_2O PDFs compare well to the observations, indicating that mixing may be reasonable in the TMS. The 60°N/S mean ages are the youngest of any CCM, and all extra-tropical mean ages are ~1.5 years too young. Surprisingly, the Antarctic N_2O profile agrees well with observations in the LS and is lower than observed in the MS - the opposite of what is expected with young air. The vortex is not isolated at any level. The fractional release curves are much steeper than the observations, meaning that even very young air has undergone significant photolysis; the fast vertical transport may explain this. (This model did not participate in the photochemical inter-comparison reported in Chapter 6.) The large fractional release explains the good agreement with the mid-latitude and vortex Cl_y time series. In the Antarctic, Cl_y at 1 hPa is about the same as it is at 50 hPa. The lack of vertical gradient is consistent with all CFC photolysis occurring at relatively low altitudes in the stratosphere. Overall, in spite of a very fast tracer-derived circulation, indications of significant vertical diffusion, and the absence of a high latitude transport barrier, compensating errors allow this model to produce realistic levels Cl_y in the Antarctic LS. The AMA grade is among the lowest of the CCMVal-2 models.

E39CA: Mean age was not available for this evaluation, however, Stenke *et al*. (2009) present mean ages at 25 hPa that are nearly a year older than E39C and close to observed values. The tape recorder phase lag is zero above the tropopause and then agrees fairly well with observations, so that the overall ascent rate is somewhat fast in the TLS. The tape recorder ascent rate is fairly close to observations throughout the TMS. There are some tracer-dependent differences in the mixing diagnostics. The tape recorder attenuation indicates very rapid mixing in both the TLS and TMS. However, the CH_4 gradients indicate relatively good mixing throughout the year in the TMS. Output for the N_2O PDF diagnostic was unavailable. The Antarctic N_2O profile is extremely high, which probably reflects the 10 hPa lid. The output for evaluating vortex isolation was unavailable, but October zonal mean meridional gradients of N_2O and CH_4 at 50 hPa are nearly flat in the SH, suggesting that this model is not likely to have a high latitude transport barrier. The mid-latitude Cl_y time series agrees well with the observations; the vortex Cl_y agrees in 1992 but is too low in 2005. Overall, the circulation is slightly fast in the TLS but good in the TMS. There is too much T-M mixing in the TLS, but the diagnostics are contradictory in the TMS. The low lid likely impacts high-latitude descent. Cl_y is reasonable in the mid-latitudes but low in high latitudes, but as this model uses an upper boundary condition for Cl_y, Cl_y cannot be used to diagnose transport.

EMAC: Mean age is not available for this model. The tape recorder phase speed indicates good agreement with the observations throughout most of the TLS and TMS. The tape recorder attenuation shows very good mixing in the TLS, but the CH_4 gradients show strong mixing in the this region. The tape recorder and CH_4 diagnostics are consistent in showing somewhat strong mixing where they overlap in the TMS, but the CH_4 indicates that the mixing is good overall from 24-32 km. The N_2O PDFs generally agree well with the observations in the SH but do not agree as well in the NH. The vortex isolation diagnostic indicates good separation at 800 K and above, but below 600 K the vortex is not strongly isolated from the mid-latitudes. Consistent with that, the Antarctic N_2O profile is too high from 30-100 hPa, indicating young air inside the vortex. Mid-latitude Cl_y is slightly lower than observations. Vortex Cl_y in 1992 is too low, and although this run did not continue to 2005,

its time series in the late 1990's shows that Cl_y has already peaked at ~2.6 ppb, more than 0.5 ppb lower than the 2005 observations. Overall, tropical ascent rates are about right, but there is too much T-M mixing in the TMS. The polar vortex barrier is too weak throughout the LS and thus the vortex lacks the necessary isolation to produce realistic Cl_y levels. Low mid-latitude Cl_y may contribute to this.

GEOSCCM: The tape recorder phase speed is variable in the TLS, but the average value compares well to observations, as does the age gradient ascent rate. Both the tape recorder and age gradient ascent rates are somewhat faster than the residual vertical velocity in the TLS, but all three agree in the TMS and are relatively close to the observed velocities. In the TLS the tape recorder attenuation indicates very good mixing rates, supported by reasonable CH_4 gradients and good mean ages. In the TMS, the CH_4 gradients are weak, which suggests too little mixing given the good ascent rates. The tropical mean age is slightly young, also indicating somewhat weak mixing given the good ascent. The N_2O PDFs show slightly too much separation above 600 K, especially in the NH. The 60°N/S mean ages agree very well with observations. The Antarctic N_2O profile agrees well in the LS but is too low above 50 hPa, suggesting descent may occur with too much isolation. The CH_4 PDFs support this, showing very low CH_4 in the vortex, persisting to below at 400 K, lower than observed. The fractional release curves and age range spanned agree very closely with the empirical curves. The mid-latitude Cl_y time series matches the observations very well and the vortex Cl_y is slightly lower than observed. Overall, the stratospheric circulation appears to be quite realistic in the tropics as well as in the high latitudes. Transport barriers may be somewhat strong in the MS. The AMA grade is the highest of the CCMVal-2 models. Although the vortex is somewhat too strong, this does not hinder its ability maintain realistic levels of Cl_y in the Antarctic spring.

LMDZrepro: The tape recorder signal could only be analysed in the TLS. The tape recorder phase speed is variable, but is overall slightly fast in the TLS. The age gradient ascent is too slow in the TLS and does not have the correct profile – the ascent is essentially constant between the tropopause and 50 hPa. The residual vertical velocity is similar to the age gradient but shows a slight minimum at ~75 hPa. In the TMS, the age gradient ascent agrees very well with both the observed tracer ascent and the residual vertical velocity. The tape recorder amplitude and CH_4 gradients indicate somewhat too much mixing in the TLS. In the TMS, where the ascent rates appear to be very good, the CH_4 profiles indicate too little mixing. The N_2O PDFs agree well with the observations but occasionally show too much separation between the tropics and extra-tropics. The 60°N/S mean ages are in good agreement with observations. The Antarctic N_2O agrees with observations from 100-20 hPa but is too low above. The vortex isolation diagnostic, which evaluates composition from 50°S-80°S, finds little mid-latitude air below 700 K, suggesting either a very large vortex or insufficient isolation. Above 700 K, strong vortex isolation is found but the isolation extends too high (2000 K), suggesting a lack of wave activity at high altitudes in this season. Fractional release curves agree very closely in slope and range with the observations. The mid-latitude Cl_y time series agrees closely with the observations. The vortex Cl_y is slightly low (by ~1σ) for both years evaluated. Overall, the unusual shape of the age gradient profile and the structure of the tape recorder water vapour anomalies below 17 km indicate problems with TLS transport. There is too much T-M mixing in the LS, but good ascent and slightly weak T-M mixing in the MS. Except for the tropical LS, mean ages are very good and the average mean age is better than most models. Antarctic descent may be reasonable, but the LS vortex appears to be too large or insufficiently isolated in spring.

MRI: The tape recorder signal could only be analysed in the TLS. The tape recorder phase speed is variable, but overall agrees well with the observations in the TLS. The age gradient ascent rates also agree well with observations and both ascent rates are close to the residual vertical velocity. In the TMS, the age gradient ascent agrees well with the observed tracer ascent up to 30 hPa but is somewhat fast above. The tape recorder attenuation indicates that TLS subtropical mixing is a bit strong, but the tropical CH_4 gradients and N_2O PDFs indicate very strong mixing in the TLS and TMS. The mean age profiles in the tropics and mid-latitudes are both in excellent agreement with the observations, yet MRI is older than all other models with good ascent rates. The 60°N/S mean ages are also in excellent agreement with observations. The Antarctic N_2O profile is slightly high in the LS but is in good agreement above. The CH_4 PDFs calculated on pressure surfaces (temperature was unavailable) have no bimodal structure whatsoever for latitudes poleward of 50°S. The CH_4 mixing ratios in the LS are characteristic of mid-, not high latitudes, in contrast with the LS N_2O which is more characteristic of vortex air. The fractional release rates of CFC-11 and CFC-12 are the same, and only the CFC-11 curve is in good agreement with the observations. The CFC-12 slope is much steeper than observed, indicating a larger release of Cl occurring at younger ages which may explain the higher than observed mid-latitude Cl_y time series. The CFC results and inconsistencies between photochemical and non-photochemical tracer diagnostics indicate that there may be problems with photolysis rates, but this model did not submit results for the photochemical inter-comparison in Chapter 6. Vortex Cl_y is much lower than observations, shows no increasing trend in the 1990's, and has very large interannual vari-

ability. The low vortex Cl_y is consistent with high CH_4 (*i.e.*, non-vortex) mixing ratios seen in the high latitude PDFs. Overall, the tropical diagnostics indicate good ascent below 30 hPa with somewhat fast ascent above. MRI has an AMA that is among the highest of all CCMs yet T-M mixing is too strong in both the TLS and TMS. This suggests that problems with ascent (fast) and mixing (strong) act to offset each other, producing good mean ages. Fractional release discrepancies, the lack of vortex isolation, and inconsistencies between indicators of high latitude composition all raise red flags.

NiwaSOCOL: There are large differences in the tropical upwelling velocities, which, along with other diagnostics, suggests excessive vertical diffusion. The tape recorder ascent rates are much faster than the observed tracer ascent rates in both the TLS and TMS. The age gradient ascent rates are also quite fast below ~40 hPa. Above 40 hPa the age gradient velocities agree well with observations. The residual vertical velocity, however, is much faster than the observed tracer-based velocities above 50 hPa. The mean age is young everywhere, consistent with rapid net vertical transport. The tape recorder signal is rapidly attenuated, and the CH_4 gradients agree well with the observations despite the fast tracer-derived ascent. Taken together, they suggest either too much T-M mixing, too much vertical diffusion, or both. Above 20 hPa, however, the CH_4 and N_2O diagnostics suggest very good T-M mixing. The 60°N/S mean ages are the 2nd youngest of any CCM, and the extra-tropical mean age is too low by at least 1 year at all levels. The Antarctic N_2O profile is much higher than observed at all levels below ~14 hPa. The vortex isolation diagnostic shows very good vortex barriers at appropriate levels from 500-1500 K, but the isolation does not persist at 450 K. Both of these diagnostics suggest too much low latitude influence in the LS vortex. While the mid-latitude Cl_y time series is in excellent agreement with observations, the vortex Cl_y is the lowest of all the CCMs, more than 1 ppb lower than observed. The Cl_y 1 hPa time series raises concerns because its maximum value is ~4 ppb, a value greater than total Cl prescribed at the surface. Overall, the model vertical transport is too fast throughout much of the tropics and there is inconsistency between the three vertical velocity estimates, suggesting excessive vertical diffusion. The polar transport barrier is too weak below 500 K. The AMA grade is among the lowest of all models. The model is unable to maintain high Cl_y levels in the vortex and Cl_y family mixing ratios are not conserved. The good agreement with mid-latitude Cl_y could result from compensating errors (Cl_y non-conservation and LS circulation issues).

SOCOL: In general, the diagnostics are very similar to NiwaSOCOL. The tape recorder and, to a lesser degree, age gradient ascent rates both indicate very rapid ascent in the TLS. In the TMS the age gradient ascent is much closer to observations than the tape recorder ascent. The mean ages are slightly older than NiwaSOCOL in the TMS, consistent with slightly slower ascent there (as diagnosed by the age gradients). As in NiwaSOCOL, the age gradient and tape recorder ascent rates show large deviations from the residual vertical velocities in the TLS, but unlike NiwaSOCOL, the age gradient agrees well with the residual vertical velocity above ~40 hPa. The tape recorder attenuation and CH_4 gradients both indicate too much mixing and/or too much vertical diffusion below 20 hPa given the fast ascent rates. As with NiwaSOCOL, the CH_4 and N_2O diagnostics indicate that T-M mixing is much better above 20 hPa. The 60°N/S mean ages are 1 year or more too young. The vortex isolation diagnostic shows very good vortex barriers at appropriate levels from 500-1500 K, but the isolation does not persist at 450 K. Mid-latitude Cl_y agrees closely with the observations. Vortex Cl_y is ~0.2 ppb higher than NiwaSOCOL but remains well below the observations and shows large interannual variability. The Cl_y 1 hPa time series also increases unrealistically to ~4 ppb. While SOCOL shows overall slightly higher mean ages than Niwa-SOCOL, this does not significantly change the transport conclusions. Overall, the model circulation is too fast in the LS, there are indications of excessive vertical diffusion, the polar transport barrier below 500 K is too weak, and the LS vortex is unable to maintain high Cl_y levels. The AMA grade is among the lowest of all models. Cl_y family mixing ratios are not conserved. The good agreement with midlatitude Cl_y could result from compensating errors (Cl_y non-conservation and LS circulation issues).

ULAQ: The tape recorder signal could only be analysed in the TLS. The phase speed is variable but shows overall good agreement with the observed tracer-derived ascent rates. The age gradient ascent is slower than observed below 60 hPa and then faster than observed from 60 to 15 hPa. The age gradient vertical velocities are somewhat faster than the residual vertical velocities, but both have the same profile, with the minimum ascent near 90 hPa instead of 50 hPa as in the observed tracer derived profile. The tape recorder attenuation indicates reasonable mixing in the TLS. The CH_4 gradients, however, are very strong in the TLS, indicating rapid mixing. In the TMS, the CH_4 profile shows good agreement with observations in JFM (indicating too much mixing given the fast ascent) but much more slowly in JAS (consistent with the ascent). The N_2O PDFs are only weakly bimodal, further suggesting too much mixing in the TMS. The 60°N/S mean ages agree very well with observations and Antarctic N_2O agrees very well with observations in the LS but is too high above 30 hPa. The vortex isolation diagnostic shows no evidence of a vortex barrier to mixing at any level. In the LS, CH_4 values are intermediate between the observed vortex and

mid-latitude mixing ratios, consistent with a lack of vortex isolation. Fractional release for CFC-11 and CFC-12 both agree well with observations although they span a slightly lower mean age range. The mid-latitude Cl_y time series agrees very well with observations but the vortex Cl_y is slightly low. Overall, the tropical ascent is somewhat slow above the tropopause and somewhat fast in the TMS, and the shape of the age gradient ascent profile is significantly different than observed. There is reasonable T-M mixing in the TLS, but it is too strong above. The AMA grade is quite high, but given the ascent rate profile and the lack of a tape recorder signal in the MS, the AMA may not be a good indicator of this model's transport credibility. The vortex either extends equator-ward of 50°S or has no transport barrier; either way, an important feature of the Antarctic stratosphere is not well-represented.

UMETRAC. A water vapour climatology is used; no tape recorder diagnostics can be evaluated. The age gradient ascent rates show very good agreement with observations. The CH_4 gradients compare well with observations in the TLS and TMS, indicating good mixing. The ages are young despite good ascent and mixing. There are no N_2O PDF diagnostics for this model. The 60°N/S mean ages are both slightly low, and, consistent with this, the Antarctic N_2O profile is a little high. The vortex isolation diagnostic could not be applied, and as the REF-B0 (time slice) experiment was evaluated, no Cl_y time series was possible. In spite of realistic tropical ascent rates, the average mean age grade is only fair, suggesting transport barriers (subtropical and/or polar) may be affecting mean age in the MS. There is insufficient output to draw more specific conclusions about the credibility of transport, particularly in the high latitudes.

UMSLIMCAT. The tape recorder signal could only be analysed in the TLS. The tape recorder phase lag is variable and is zero (*i.e.*, infinite phase speed) over a portion of the TLS, so that the average ascent is considerably faster than observed. The age gradient ascent, however, agrees fairly well with the observations in the TLS. In the TMS, the age gradient ascent becomes progressively faster than observed. The tape recorder attenuation and CH_4 gradients indicate somewhat too much mixing in the TLS. TMS CH_4 gradients match the observations fairly well, indicating too much mixing given the fast ascent rates. The fact that the tropical mean ages are relatively good while the mid-latitude mean ages are too young also suggests rapid ascent with too much in-mixing of mid-latitude air in the TMS. The N_2O PDFs show weak bimodality in the LS but somewhat stronger bimodality in the MS. The 60°N mean age is very good and the 60°S age is slightly low. The Antarctic N_2O profile is realistic in the LS and MS, and the vortex isolation diagnostic shows excellent vortex barriers at the appropriate levels. Both the mid-latitude and vortex Cl_y agree very well with observations, however, CFC-11 and CFC-12 J-values are well outside the expected range (see Chapter 6). Overall, this model has a fast circulation in the TMS along with too much T-M mixing throughout the TLS and TMS. As these attributes have opposite effects on mean age, the AMA is fairly high. Vortex isolation and Cl_y are realistic, but the CFC J-values are significantly different from other models.

UMUKCA-METO. A water vapour climatology is used; no tape recorder diagnostics can be evaluated. The age gradient ascent rates are much slower than observed in the TLS. They are also slower than observed in the TMS, but the difference between the observations and the model is smaller. The UMUCKA models are the only models for which the age gradient ascent rates are considerably slower than the residual vertical velocity throughout the tropical stratosphere. The mean ages are very old, consistent with the slow age-derived circulation. The shape of the tropical age profile, the CH_4 gradients, and the N_2O PDFs all indicate too little mixing in the TMS. The 60°N/S mean ages are 2 or more years too old. This is consistent with the Antarctic N_2O profile, which is much lower than observed. The CH_4 PDFs show a very strong vortex extending to higher than observed levels (at least 2000 K). The PDFs also show that the vortex has no influence from lower latitude air down to ~550 K while observations indicate the influence of lower latitude air at 700 K and below. The fractional release curves are less steep than observed, meaning that for a given mean age they have released less Cl than expected. This is consistent with the very slow circulation: slow ascent allows old mean ages to occur at lower altitudes where photolysis rates are lower. Both the vortex and mid-latitude Cl_y time series are much higher than observed. There is a known bug in this simulation related to wet deposition of HCl, and this has been reported to increase Cl_y levels in the stratosphere. Overall, this model has a very slow circulation leading to very old mean ages throughout the stratosphere. The AMA grade is among the lowest of all models. The Antarctic vortex is too strong in the LS and in the US/lower mesosphere.

UMUKCA-UCAM. A water vapour climatology is used; no tape recorder diagnostics can be evaluated. Dynamics are reported to be identical to UMUKCA-METO. The diagnostics for this model are very similar to those for UMUCKA-METO, and indicate a very slow age-derived circulation in the TLS with somewhat better agreement with observations in the TMS. UMUCKA-UCAM has slightly slower age-based ascent and slightly older (0.3-0.4 years) tropical and mid-latitude mean age profiles than UMUCKA-METO. As in UMUCKA-METO, the tropical age profile, CH_4 gradients, and N_2O PDFs indicate too lit-

tle mixing in the TMS. The Antarctic N_2O behaviour is the same, and although output was not available to calculate the CH_4 PDFs, there is no reason to expect that vortex behaviour would be different from UMUKCA-METO. The mid-latitude Cl_y time series is higher than observed and the vortex Cl_y is in good agreement. The high mid-latitude Cl_y could be related to the very old mean ages found in the mid-latitude LS. Overall, this model has a very slow circulation, leading to very old mean ages throughout the stratosphere. The AMA grade is the lowest of all models. The Antarctic vortex is probably too strong in the LS and above.

WACCM. The tape recorder phase speed is variable but overall shows quite good agreement with observations in both the TLS and TMS. The age gradient vertical velocities also show good agreement with the observations, and both the age gradient and the tape recorder ascent are relatively close to the residual vertical velocity. The tape recorder attenuation indicates somewhat too much mixing in the TLS, but slightly weak mixing in the TMS. The CH_4 gradients also indicate too much mixing in the TLS and too little in the TMS. The N_2O PDFs indicate too much tropical isolation in some seasons. The tropical mean age profile is quite good in the LS and MS, but the mid-latitude mean age profile is too young above 50 hPa, consistent with too little T-M mixing in the MS. The 60°N/S mean ages are in very good agreement with observations. The Antarctic N_2O profile has overall good agreement with observations in the LS and MS, and the vortex isolation in the LS has excellent agreement with observations, including the appearance of lower latitude influence in vortex air at 700 K and below. However, a strong vortex extends all the way to the lower mesosphere instead of ending at ~1400 K. Fractional release curves have steeper slopes than observed, especially for CFC-12. This means more Cl is released for a given

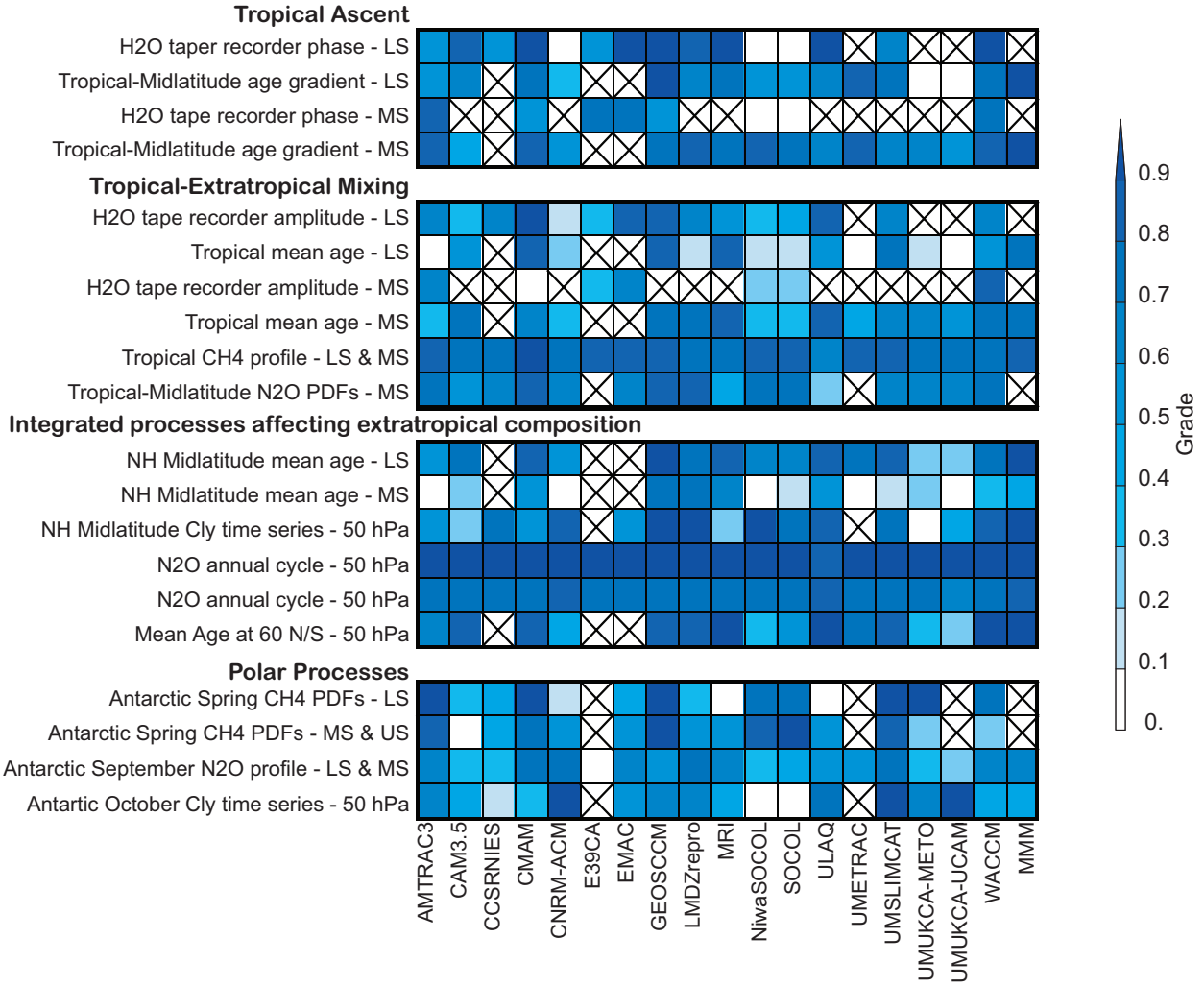

Figure 5.19: Quantitative assessment of model performance on transport diagnostics. The colour bar indicates the grade on each diagnostic test discussed in the chapter, with deeper blue indicating better agreement with observations. 'X' indicates no diagnostic could be calculated. LS, MS, and US refer to the lower stratosphere, middle stratosphere, and upper stratosphere, respectively.

mean age than expected, although CFC-11 and CFC-12 J-values are in the expected range (see Chapter 6). The mid-latitude Cl_y time series has excellent agreement with observations but vortex Cl_y is lower than observed for reasons that are not clear. Overall, the circulation is fairly realistic, with minor T-M mixing issues, and the LS polar transport barrier is very good. The AMA grade is fairly high.

5.5.2 Overall CCMVal-2 Model Transport Summary

A summary of the quantitative evaluation of the transport diagnostics is given in **Figure 5.19**. There are significant problems with simulation of the tropical stratosphere in the CCMVal-2 models. Tropical ascent and mixing across the subtropics are crucial to distributing ozone depleting substances in the stratosphere, and these transport deficiencies affect modelled abundances of Cl_y. Of the 12 models with both tape recorder and age output, only four score higher than 0.7 (*i.e.,* are within ~1σ of the observations) on both measures of tropical ascent in the tropical lower stratosphere. Of the remaining eight models, six have faster-than-observed ascent according to at least one diagnostic and two fail to reproduce the shape of the observed profile in the age gradient. Of the six models with only one available ascent diagnostic, two score higher than 0.7, two show fast tracer-derived ascent relative to observations, and two have slow ascent. The performance in the middle stratosphere is somewhat better, with eight models showing improvement in at least one ascent diagnostic. However, the fraction of models that score better than 0.7 on both diagnostics is the same as in the lower stratosphere: 1/3. The models that perform well on both tracer ascent diagnostics also tend to show good consistency between their tracer-derived upwelling and the tropical residual vertical velocity, $\overline{w}^*$. Lack of consistency between these measures of ascent may indicate numerical errors in transport or problems with vertical diffusion, horizontal or vertical eddy tracer fluxes, or phasing between temporal variability in tracers and in the circulation.

The tape recorder amplitude is the only tropical-extra-tropical mixing diagnostic that is fully independent of the circulation, but the other mixing diagnostics are independent of one another and are affected by the circulation in different ways. Thus, the performance relative to the suite of diagnostics as a whole is a more accurate assessment of subtropical mixing than performance on any single mixing diagnostic. The range of model performance here is smaller than the range on the tropical ascent diagnostics: no model has an average mixing grade higher than 0.8 or lower than 0.4 (the range for the average of the ascent diagnostics, on the other hand, is 0.15 to 1.0). Only five models have an average mixing score higher than 0.7 (all of them also performed well on the ascent diagnostics). The remaining models all have too much tropical-extra-tropical mixing and/or too much vertical diffusion in the lower stratosphere.

Excessive subtropical mixing would tend to increase the mean age in the models and is a likely explanation for the fact that most models have good mean ages in the tropical lower stratosphere despite having relatively fast circulations. However, a rapid circulation coupled with an inability to maintain tropical isolation will lead to inadequate conversion of CFCs to Cl_y. In the middle stratosphere, six models had at least some indication of too little mixing. This may contribute to the young modelled mean ages in the tropical middle stratosphere, though the error in mean age at a given level depends on the integrated errors in transport below that level and thus the relative dependence on local processes decreases with height above the tropopause.

The mean age is a sensitive function of both the circulation and mixing and the distribution of mean ages throughout the stratosphere reflects the balance between these two transport processes and their variations with height. The average of mean age grades evaluated at a wide range of latitudes and altitudes can be used to assess a model's overall transport fidelity. While it is theoretically possible to achieve the correct mean age everywhere through compensating errors in ascent and quasi-horizontal mixing, in practice it is unlikely to happen because it would require the two to perfectly offset each other throughout the stratosphere. The average mean age (AMA) grade should give a more reliable indication of transport credibility than individual mean age diagnostics and should be evaluated in addition to the individual mean age diagnostics.

The AMA metric is calculated from seven mean age grades: tropical LS (90-50 hPa), tropical MS (30-10 hPa), NH mid-latitude LS (90-50 hPa), NH mid-latitude MS (30-10 hPa), SH mid-latitude LS (50 hPa), 60°N (50 hPa), and 60°S (50 hPa). **Figure 5.20** illustrates the relationship between the AMA metric and several key diagnostic quantities: tropical LS ascent, tropical-mid-latitude mixing, Antarctic descent, and LS vortex isolation. Tropical LS ascent is the average of the two independent diagnostics for ascent, and the tropical-mid-latitude mixing quantity plotted is the average over all mixing grades available for each model (as many as six). (See Table 5.1 for a complete list of diagnostics.) The AMA shows a positive correlation with all the key diagnostics except for vortex isolation. The grey-shaded area shows models that score greater than 0.6 for the diagnostics plotted.

The relationship between a model's AMA and mean LS ascent grade is particularly powerful (Figure 5.20). Models falling in the grey-shaded upper right quadrant have successfully simulated tropical LS ascent and mean age at all locations, implying that tropical-mid-latitude

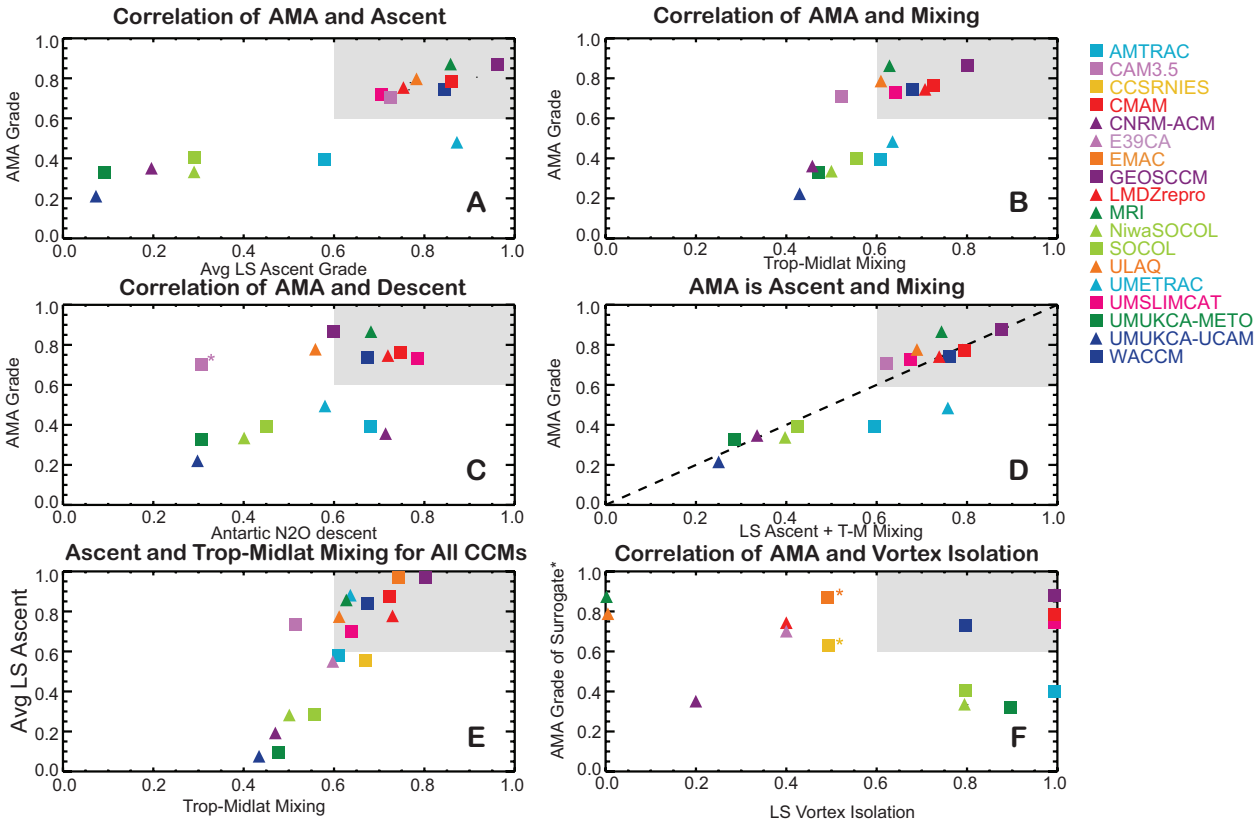

Figure 5.20: Correlations between the average mean age (AMA) grade and four fundamental diagnostic quantities: tropical LS ascent (A), tropical-mid-latitude mixing (B), vortex descent (C), and LS vortex isolation (F). The grey shaded area identifies realistic model performance for the metrics shown in that panel. The AMA is calculated from 7 mean age diagnostic scores (see text). The average LS ascent grade (panel A) is the average of the tape recorder and mean age gradient tropical ascent grades in the LS. The tropical-extra-tropical mixing grade (panel B) is the average of all tropical-extra-tropical mixing grades (Section 2 of Figure 5.19). The vortex descent grade (panel C) is taken from the Antarctic N_2O profiles. CAM3.5 has a lid of ~3 hPa that affects vortex descent; this is noted by an asterisk in panel C. Panel D shows that the AMA is a strong function of both tropical LS ascent and tropical-mid-latitude mixing. Because of this relationship, a surrogate for AMA can be calculated for models that did not submit age of air output. This allows their vortex isolation behaviour to be plotted in panel F; the models with AMA surrogate are labelled with an asterisk. Panel E shows the tropical ascent and tropical-mid-latitude mixing performance for all CCMs participating in this evaluation.

mixing is probably good as well. (This is borne out by Figure 5.20b which shows that nearly the same set of models performs well on the mixing diagnostics.) A model that falls in the lower right quadrant has successfully simulated the tropical ascent, but not mean age; thus it must have a good circulation but too much or too little mixing across the subtropics. UMETRAC is the only model that falls in the lower right quadrant. Unfortunately, it does not have output for the tape recorder or N_2O diagnostics, so it is difficult to draw conclusions about its mixing, but the CH_4 gradients do independently suggest that it may have too little mixing in the TMS in some seasons. The upper left quadrant indicates a poor circulation and incorrect mixing; the mixing would have to be compensating for the circulation to give reasonable ages. The lower left quadrant indicates problems with tropical ascent, but nothing can be inferred about the mixing across the subtropics since the circulation alone could be responsible for the poor mean age grades. However, note that the models in the lower quadrants do, in fact, have low mean values on the tropical-extra-tropical mixing diagnostics, as seen in Figure 5.20b. This correlation between the ascent and mixing diagnostics is not surprising, given that both depend on the strength and distribution of wave activity in the extratropics. Figure 5.20c shows the relationship between polar descent and the AMA. Tropical ascent and polar descent are physically linked through the diabatic circulation, thus most of the models that do well in Figure 5.20a also do well in Figure 5.20c (except CAM3.5, which has a low lid and cannot fully simulate Antarctic descent). Because there

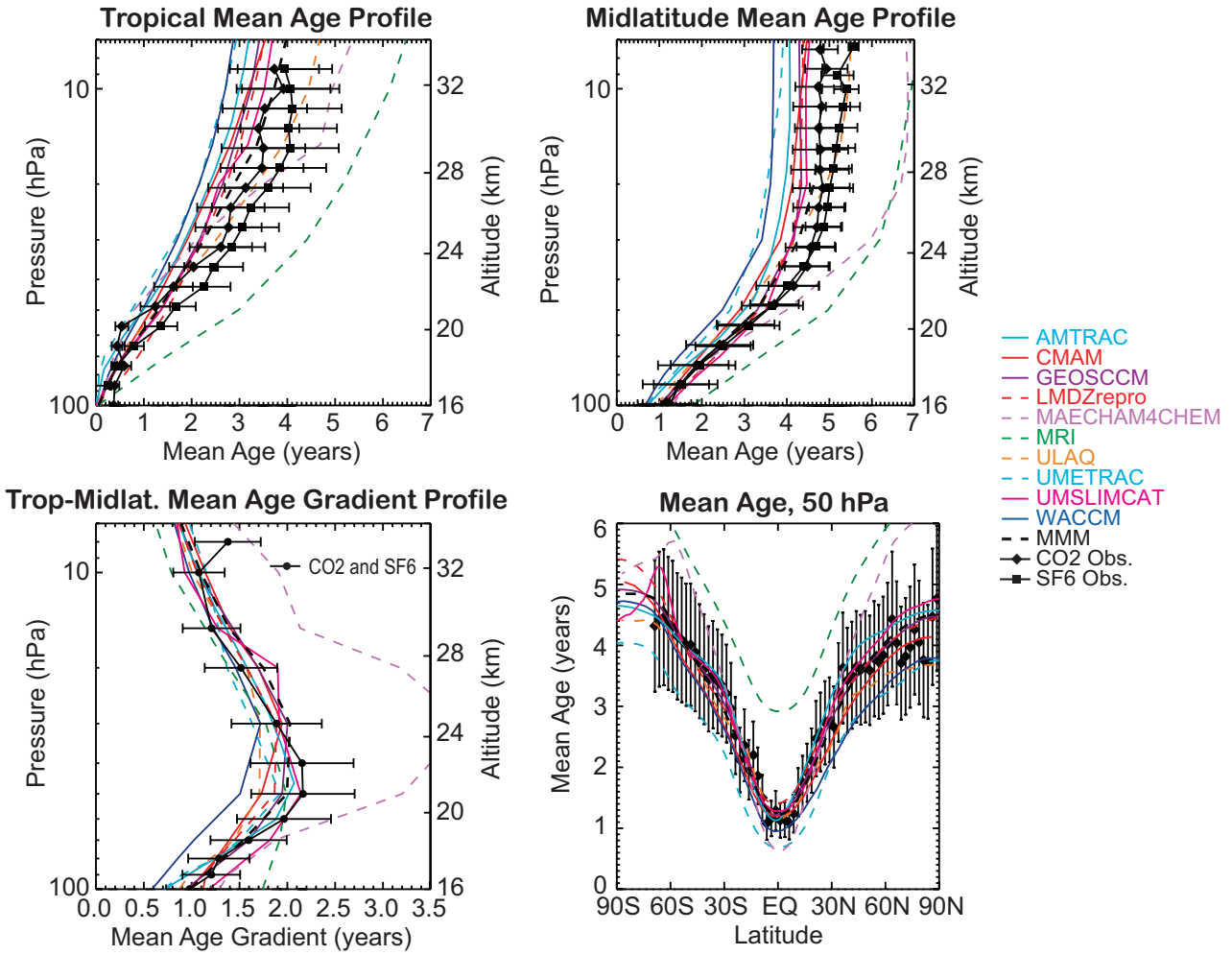

Figure 5.21: Mean age from 10 CCMs participating in CCMVal-1 and their multi-model mean (MMM). Black symbols are the observed mean age profiles derived from CO_2 (diamonds) and SF_6 (squares) for the tropics (10°N -10°S) and mid-latitudes (35°N–45°N), and the latitudinal distribution of mean age at 50 hPa ; the same as in Figure 5.5. The bottom left panel shows the difference between the average of the observed tropical profiles and mid-latitude profiles on pressure surfaces. All uncertainties shown are 1σ.

is a physical basis for the correlations shown, evaluation of these fundamental diagnostics, including the AMA, is essential for building model credibility.

While the AMA grade is a very useful quantity for evaluating a model's overall transport fidelity, the age of air output necessary for this calculation was not available from all models. Nevertheless, the strong correlations between AMA, LS ascent, and tropical-mid-latitude mixing allow a surrogate for AMA to be calculated for the models without age of air (CCSRNIES, E39CA, and EMAC). Figure 5.20d shows a roughly 1:1 relationship between the mean of the average LS ascent and average T-M mixing grades and the AMA grade for the 15 models with age output. Figure 5.20e shows the relationship between LS ascent and tropical-mid-latitude mixing for all 18 CCMs participating in this evaluation. The three models that have no AMA grade (CCSRNIES, E39CA, and EMAC) fall along the same line as the other CCMs, suggesting that the average of their LS ascent and T-M mixing grades can be used as a surrogate for AMA. This plot shows that EMAC, which could not be evaluated using the AMA, also performs very well on the ascent and mixing diagnostics.

Figure 5.20f shows correlation between LS vortex isolation and the AMA (or its surrogate) for all 18 models. The lack of correlation suggests that the transport barrier at the vortex edge is to first order independent of the overall transport circulation. Because it is not correlated with other key diagnostics and because it is a required feature for producing a realistic ozone hole, it is important to include this transport diagnostic for simulations predicting future ozone.

Figure 5.20 shows that nine of the 18 CCMs perform acceptably (> 0.6) on the AMA metric or its surrogate (CAM3.5, CMAM, EMAC, GEOSCCM, LMDZrepro,

MRI, ULAQ, UMSLIMCAT, and WACCM), and eight of these also have reasonable tropical ascent and tropical-midlatitude mixing (all but CAM3.5). Of these eight models, six also perform well on the diagnostic of Antarctic descent (models CMAM, GEOSCCM, LMDZrepro, MRI, ULAQ, and WACCM). Of these six, LMDZrepro and MRI do not create an isolated vortex in the Antarctic lower stratosphere. Only CMAM, GEOSCCM, UMSLIMCAT, and WACCM, demonstrate credible transport performance in all key areas.

Grewe and Sausen (2009) proposed that the grading metric and observational data sets used in WE08 are unlikely to be able to distinguish between realistic and unrealistic model behaviour due to uncertainties and interannual variability in the data sets; that is, they believe there is a high potential for the grades to be meaningless. Figure 5.20 demonstrates the overall credibility of the diagnostics, as well as the WE08 grading metric used, by showing strong correlations between the grades for average mean age and the average of the tropical LS ascent and tropical-mid-latitude mixing grades. If the individual grades that make up the combined grades were meaningless, that is, if a grade of 1 were not statistically different from a grade of 0, the average of those grades would be a random number bearing no relation to model transport behaviour. In that case, plotting those metrics against each other (Figures 5.20 a-e) would produce random scatter. On the contrary, Figure 5.20 shows a remarkable degree of correlation between the grades for these fundamental transport diagnostics. The relationship between these diagnostics demonstrates that the observations used to create them are physically meaningful and have statistically significant information, and that the grading metric itself is not a random, statistically insignificant quantity.

5.5.3 Summary of 21st century transport changes

REF-B2 output was analysed from 10 CCMs. Despite the spread in model performance revealed by the transport diagnostics, all of these models predict a faster circulation and younger mean age at the end of the 21st century, indicating that this is a robust result that depends on large-scale forcing. Only six of these models had water vapour anomalies with a clearly identifiable tape recorder signal in the TLS (CAM3.5, CCSRNIES, CMAM, LMDZrepro, SOCOL, and WACCM), and only three had a tape recorder that persisted in the TMS (CMAM, SOCOL, and WACCM). The changes in the tape recorder ascent rate provide a less clear picture of changes in the circulation than the age gradient, but all models but one (SOCOL) predict an increase in the phase speed over most of the tape recorder altitude range. The base of the tape recorder is higher at the end of the 21st century than during present day in all but one model. The models also generally predict less rapid attenuation of the tape recorder signal in the future, which is consistent with the predicted changes in circulation.

Comparisons of the future Antarctic vortex predicted by 5 CCMs showed no agreement on whether the vortex would be larger or smaller in the future. All showed a slightly deeper minimum in the LS/MS PDFs (*i.e.*, a stronger barrier) and all showed some indication of increased horizontal mixing in the upper stratosphere.

5.5.4 Comparison to CCMVal-1 model transport

Mean age can be used to examine differences in transport performance between CCMVal-1 and CCMVal-2. **Figure 5.21** shows tropical and mid-latitude mean age profiles, the mean age gradient, and 50 hPa mean age for CCMVal-1 models in the same format as the CCMVal-2 models shown in Figure 5.5. The multi-model mean (MMM) ascent rate as assessed by the tropical-mid-latitude age gradients is nearly identical for both sets of models, but the spread of the CCMVal-2 models is much greater. At 50 hPa, the age gradients of only 2 of 10 CCMVal-1 models fall outside the observational uncertainty while 7 of 15 CCMVal-2 models do. The same is true for all 50 hPa mean ages. CCMVal-1 models are tightly clustered with only 2 models frequently outside the uncertainties and the MMM agrees very closely with the observations. The CCMVal-2 MMM is in almost as good agreement in spite of a much larger spread. This is largely due to a fortuitous balance of models with very old age cancelling the effects of models with very young age.

In contrast to the age diagnostics, the CCMVal-2 tape recorder phase speed shows improvement in both the difference between the MMM and observations and in the spread of model performance relative to CCMVal-1. This at first seems contradictory, but of the seven models with both diagnostics available for both CCMVal-1 and CCMVal-2, six (CMAM, GEOSCCM, WACCM, LMDZrepro, MRI, and ULAQ) scored better than 0.7 on at least one tropical ascent diagnostic in this assessment. The changes between CCMVal-1 and CCMVal-2 varied for these six models, and also varied for the two tropical ascent diagnostics from a single model, but all performed relatively well on these diagnostics in both assessments. (The tape recorder performance was only considered for the TLS for LMDZrepro, MRI, and ULAQ.) Four of the five CCMVal-2 models with the worst performance on the tropical ascent diagnostics did not participate in CCMVal-1 (NiwaSOCOL, UMUCKA-METO, UMUCKA-UCAM, and CNRM-ACM).

The mid-latitude mean age profile comparison reveals some interesting differences between the sets of

models. While both sets of models have young and old outliers, most of the CCMVal-1 models form a fairly compact cluster underneath the observational uncertainties. The CCMVal-2 models form a looser cluster, and relatively few fall under the observations above 30 hPa (25 km); overall they are about 0.5 years younger than CCMVal-1 models in the middle stratosphere. Although the tropical mean profiles show the same spread for both sets of models, the CCMVal-2 MMM mean age is younger than the CCMVal-1 MMM and does not agree as well with the observations.

It is not clear that there has been any improvement in performance among the models that participated in both CCMVal-1 and CCMVal-2. Two of the best-performing models from this assessment showed better agreement with the observations in CCMVal-1 (GEOSCCM and CMAM), and the other models show mixed results depending on the diagnostic. However, eight of the 11 models that participated in both assessments appear in the upper-right corner of Figure 5.20e, indicating reasonable performance on the tropical ascent and mixing diagnostics. The addition of four new models, all appearing in the lower left corner of Figure 5.20e, accounts for a good deal of the increase in model spread between Figure 5.21 and Figure 5.5. It is only the fortuitous cancellation of the models that perform poorly that gives a CCMVal-2 MMM in good agreement with the observations. This fact reinforces the importance of the goals of this report, namely, to reduce uncertainties in model predictions by using observationally-based diagnostics to understand model performance and determine model credibility. A multi-model mean calculated only from models with proven transport credibility would represent a step forward in reducing uncertainties in chemistry climate model predictions.

5.5.5 Requirements for transport credibility

Realistic representation of several key aspects of transport should be considered essential for credibility. They are

1. local conservation of chemical family mixing ratios (*e.g.*, Cl_y),
2. realistic tropical ascent in the LS,
3. realistic mixing between the tropics and extra-tropics in the LS and MS,
4. close agreement with *all* mean age diagnostics, that is, a high score for the average mean age grade, and
5. generation of an isolated lower stratospheric Antarctic vortex.

All of these aspects of transport are necessary for the simulation of realistic levels of vortex Cl_y. Models that reasonably represent these essential physical processes have demonstrated the credibility necessary for prediction of future stratospheric composition.

The evaluations presented in this chapter indicate transport improvement efforts in CCMs should concentrate on the simulation of the tropical lower stratosphere. Improvements are needed in the ascent rate profile below 50 hPa in the tropics and in the rate of mixing between the tropics and mid-latitudes, which is currently too strong in most models. In addition, discrepancies between a model's residual vertical velocity in the tropics and its tracer derived velocities suggests possible problems with vertical diffusion or numerics.

References

Andrews, D.G., Holton, J.R., and Leovy, C.B, 1987. Middle Atmospheric Dynamics, 489 pp. Academic, San Diego, CA.

Andrews, A. E., K. A. Boering, B. C. Daube, S. C. Wofsy, M. Loewenstein, H. Jost, J. R. Podolske, C. R. Webster, R. L. Herman, D. C. Scott, G. J. Flesch, E. J. Moyer, J. W. Elkins, G. S. Dutton, D. F. Hurst, F. L. Moore, E. A. Ray, P. A. Romashkin, and S. E. Strahan, 2001. Mean ages of stratospheric air derived from in situ observations of CO_2, CH_4, and N_2O, *J. Geophys. Res.*, **106**, 32,295–32,314.

Boering, K. A., Wofsy, S. C., Daube, B. C., Schneider, H. R., Loewenstein, M., Podolske, J. R., and Conway, T. J., 1996. Stratospheric Mean Ages and Transport Rates from Observations of Carbon Dioxide and Nitrous Oxide, *Science,* **274**, doi: 10.1126/science.274.5291.1340.

Douglass, A. R. Schoeberl, M. R., Stolarski, R. S., Waters, J. W., Russell, J. M., Roche, A. E., and Massie, S.T., 1995. Interhemispheric differences in springtime production of HCl and $ClONO_2$ in the polar vortices, *J. Geophys. Res.*, **100**, 13,967-13,978.

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

Douglass, A. R., Stolarski, R.S., Schoeberl, M.R., Jackman, C.H., Gupta, M., Newman, P.A., Nielsen, J.E., and Fleming, E., 2008. Relationship of loss, mean age of air and the distribution of CFCs to stratospheric circulation and implications for atmospheric lifetimes, *J. Geophys. Res.*, **113**, doi:10.1029/2007JD009575.

Engel, A., T. Möbius, H. Bönisch, U. Schmidt, R. Heinz, I.

Levin, E. Atlas, S. Aoki, T. Nakazawa, S. Sugawara, F. Moore, D. Hurst, J. Elkins, S. Schauffler, A. Andrews, and K. Boering, 2009. Age of stratospheric air unchanged within uncertainties over the past 30 years, *Nat. Geosci.*, **2**, 28-31.

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.

Garcia, R. R., and Randel, W. J., 2008. Acceleration of the Brewer-Dobson circulation due to increases in greenhouse gases, *J. Atmos. Sci.*, **65**, 2731-2739, doi:10.1175/2008JAS2712.1.

Glatthor, N., T. von Clarmann, H. Fischer, B. Funke, U. Grabowski, M. Höpfner, S. Kellmann, M. Kiefer, A. Linden, M. Milz, T. Steck, G. P. Stiller, G. Mengistu Tsidu, and D-Y. Wang, 2005. Mixing processes during the Antarctic vortex split in September-October 2002 as inferred from source gas and ozone distributions from ENVISAT-MIPAS, *J. Atmos. Sci.*, **62**, 787-800.

Gray, L.J. and Russell, J.M., 1999. Interannual variability of trace gases in the subtropical winter stratosphere, *J. Atmos. Sci.*, **56**, 977-993.

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

Hall, T. M., 2000. Path histories and time scales in stratospheric transport: Analysis of an idealized model, *J. Geophys. Res.*, **105**, 22,811-22,823.

Hall, T. M., and R. A. Plumb, 1994. Age as a diagnostic of stratospheric transport, *J. Geophys. Res.*, **99**, 1059-1070, doi:10.1029/93JD03192.

Hall, T.M., Waugh, D.W., Boering, K.A., and Plumb, R.A. (1999), Evaluation of transport in stratospheric models, *J. Geophys. Res.*, **104**, 18815-18839.

Lambert, A., W. G. Read, N. J. Livesey, M. L. Santee, G. L. Manney, L. Froidevaux, D. L. Wu, M. J. Schwartz, H. C. Pumphrey, C. Jimenez, G. E. Nedoluha, R. E. Cofield, D. T. Cuddy, W. H. Daffer, B. J. Drouin, R. A. Fuller, R. F. Jarnot, B. W. Knosp, H. M. Pickett, V. S. Perun, W. V. Snyder, P. C. Stek, R. P. Thurstans, P. A. Wagner, J. W. Waters, K. W. Jucks, G. C. Toon, R. A. Stachnik, P. F. Bernath, C. D. Boone, K. A. Walker, J. Urban, D. Murtagh, J. W. Elkins, and E. Atlas, 2007. Validation of the Aura Microwave Limb Sounder middle atmosphere water vapor and nitrous oxide measurements, *J. Geophys. Res.*, **112**, doi:10.1029/2007JD008724.

Lary, D. J., Waugh, D. W., Douglass, A. R., Stolarski, R. S., Newman, P. A., and Mussa, H., 2007. Variations in stratospheric inorganic chlorine between 1991 and 2006, *Geophys. Res. Lett.*, **34**, doi:10.1029/2007GL030053.

Li, F., Austin, J., and Wilson, J., 2008. The strength of the Brewer-Dobson circulation in a changing climate: Coupled chemistry-climate model simulations, *J. Clim.*, **21**, 40-57, doi:10.1175/2007JCLI1663.1.

Livesey, N.J., et al., (2007). Earth Observing System (EOS) Aura Microwave Limb Sounder (MLS) Version 2.2 Level 2 data quality and description document, JPL D-33509, 2007.

McLandress, C. and Shepherd, T. G., 2009. Simulated anthropogenic changes in the Brewer-Dobson circulation, including its extension to high latitudes, *J. Clim.*, **22**, 1516-1540, doi:10.1175/2008JCLI2679.1.

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

Mote, P. W., Dunkerton, T., McIntyre, M., Ray, E., Haynes, P., and Russell, J., 1998. Vertical velocity, vertical diffusion, and dilution by midlatitude air in the tropical lower stratosphere, *J. Geophys. Res.*, **103**, 8651-8666.

Neu, J. L. and Plumb, R. A., 1999. Age of air in a 'leaky pipe' model of stratospheric transport, *J. Geophys. Res.*, **104**, 19,243-19,255.

Neu, J. L, Sparling, L.C., and Plumb, R.A. (2003), Vari-

ability of the subtropical 'edges' in the stratosphere, *J. Geophys. Res.*, **108**, doi:10.1029/2002JD002706.

Oman, L., Waugh, D. W., Pawson, S., Stolarski, R. S., and Newman, P. A., 2009. On the influence of anthropogenic forcings on changes in the stratospheric mean age, *J. Geophys. Res.*, **114**, doi:10.1029/2008JD010378.

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

Read, W., A. Lambert, J. Bacmeister, R. E. Cofield, L. E. Christensen, D. T. Cuddy, W. H. Daffer, B. J. Drouin, E. Fetzer, L. Froidevaux, R. Fuller, R. Herman, R. F. Jarnot, J. H. Jiang, Y. B. Jiang, K. Kelly, B. W. Knosp, L. J. Kovalenko, N. J. Livesey, H.-C. Liu, G. L. Manney, H. M. Pickett, H. C. Pumphrey, K. H. Rosenlof, X. Sabounchi, M. L. Santee, M. J. Schwartz, W. V. Snyder, P. C. Stek, H. Su, L. L. Takacs, R. P. Thurstans, H. Vömel, P. A. Wagner, J. W. Waters, C. R. Webster, E. M. Weinstock, and D. L. Wu, 2007. Aura Microwave Limb Sounder upper tropospheric and lower stratospheric H_2O and relative humidity with respect to ice validation, *J. Geophys. Res.*, **112**, doi:10.1029/2007JD008752.

Rosenlof, K., 1995. Seasonal cycle of the residual mean meridional circulation in the stratosphere, *J. Geophys. Res.*, **100**, 5173-5191.

Santee, M. L., Froidevaux, L., Manney, G. L., Read, W. G., Waters, J. W., Chipperfield, M. P., Roche, A. E., Kumer, J. B., Mergenthaler, J. L., and Russell III, J. M., 1996. Chlorine deactivation in the lower stratospheric polar regions during late winter: Results from UARS, *J. Geophys. Res.*, **101**, 18,835-18,859.

Schauffler, S.M., E. L. Atlas, S. G. Donnelly, A. Andrews, S. A. Montzka, J. W. Elkins, D. F. Hurst, P. A. Romashkin, G. S. Dutton, V. Stroud, 2003. Chlorine budget and partitioning during the Stratospheric Aerosol and Gas Experiment (SAGE) III Ozone Loss and Validation Experiment (SOLVE), *J. Geophys. Res.*, **108**, doi:10.1029/2001JD002040.

Sparling, L.C., 2000. Statistical perspectives on stratospheric transport, *Rev. Geophys.*, **38**, 417-436.

Schoeberl, M.R., Douglass, A.R., Stolarski, R.S., Pawson, S., Strahan, S., and Read, W., 2008 Comparison of lower stratospheric tropical mean vertical velocities, *J. Geophys. Res.*, **113**, doi:10.1029/2008JD010221.

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

Stiller, G.P., T. von Clarmann, M. Höpfner, N. Glatthor, U. Grabowski, S. Kellmann, A. Kleinert, A. Linden, M. Milz, T. Reddmann, T. Steck, H. Fischer, B. Funke, M. López-Puertas, and A. Engel, 2008. Global distribution of mean age of stratospheric air from MIPAS SF_6 measurements, *Atmos. Chem. Phys.*, **8**, 677-695.

Strahan, S. E. and Douglass, A. R., 2004. Evaluating the credibility of transport processes in simulations of ozone recovery using the Global Modeling Initiative three-dimensional model, *J. Geophys. Res.*, **109**, doi:10.1029/2003JD004238.

Strahan, S. E., and Polansky, B. C., 2006. Meteorological implementation issues in chemistry and transport models, *Atmos. Chem. Phys.*, **6**, 2895-2910.

Waugh, D., 2009. Atmospheric dynamics: The age of stratospheric air, *Nature Geo.*, **2**, 14-16, doi:10.1038/ngeo397.

Waugh, D. W. and Hall, T. M., 2002. Age of stratospheric air: Theory, observations, and models, *Rev. Geophys.*, **40**, doi:10.1029/2000RG000101.

Waugh, D.W., and Eyring, V., 2008. Quantifying performance metrics for stratospheric-resolving chemistry-climate models, *Atmos. Chem. Phys.*, **8**, 5699-5713.

World Meteorological Organization (WMO)/United Nations Environment Programme (UNEP), 2007. *Scientific Assessment of Ozone Depletion: 2006*, World Meteorological Organization, Global Ozone Research and Monitoring Project, Report No. 50, Geneva, Switzerland.