Issues in Stratosphere-troposphere Coupling

Theodore G. SHEPHERD
Department of Physics, University of Toronto, Toronto, Canada

(Manuscript received 2 August 2001, in revised form 17 March 2002)

Abstract

An overview is given of current issues concerning the coupling between the stratosphere and troposphere. The tropopause region, more generally the upper troposphere/lower stratosphere, is the region of direct contact where exchange of material takes place. Dynamical coupling through angular momentum transfer by waves occurs nonlocally, and provides a generally negative torque on the stratosphere which drives an equator-to-pole circulation (i.e., towards the Earth's axis of rotation). This wave-driven circulation is the principal mechanism for intraseasonal and interannual variability in the extratropical stratosphere. Although such variability is generally dynamical in origin, there are important chemical and radiative feedbacks. The location of the tropopause has implications for radiative forcing of climate, through its effect on the distribution of relatively short-lived greenhouse gases (ozone and water vapour). Some outstanding puzzles in our current understanding are identified. Attention is focused on possible climate sensitivities, and how these may be tested and constrained. Results from the Canadian Middle Atmosphere Model (CMAM), a fully interactive radiative-chemical-dynamical general circulation model, are used to illustrate some of the points.

1. The stratosphere and troposphere

The very term “stratosphere-troposphere coupling” begs the question of what the stratosphere and troposphere are, or, more specifically, what the distinction is between them. The traditional meteorological distinction is in terms of lapse rate: the troposphere is weakly stratified (to vertical displacements), the stratosphere strongly stratified (Fig. 1). This contrast reflects the nature of the operative radiative-dynamical balances. Solar heating of the Earth’s surface leads to a radiative equilibrium state that is dynamically unstable, either convectively (as in the tropics) or baroclinically (as in the extratropics). The heat transfer from the resulting large-scale motion, both vertical and meridional, occurs over a region of finite depth that we may consider to be the troposphere. Within this region, transport timescales are relatively fast, ranging from hours for convective transport to days for baroclinic transport. In fact an interesting alternative definition of the troposphere is the region wherein air parcels have been in contact with the surface layer within a certain time period such as one week (Esler et al. 2002). This strong connection with the surface layer is an important characteristic of the troposphere, with significant implications for climate.

Above this region of large-scale turbulent heat transfer, the atmosphere is comparatively quiescent. To a first approximation, the radiative equilibrium state $T_{\text{rad}}$ is dynamically stable and departures from this state occur only through external forcing by waves propagating up from the troposphere. Important exceptions are effects of time dependence through the seasonal cycle in $T_{\text{rad}}$ (Garcia 1987), and the mid-
Atmospheric waves transfer angular momentum and energy (but not heat) from the surface of the Earth and the troposphere into the region above. Their contribution to the energy budget is negligible below about 80 km, being dwarfed by radiative processes, but their impact on the angular momentum budget is considerable; the wave forcing of this region is essentially mechanical. The torque associated with angular momentum transfer by waves is called wave drag. In the stratosphere, the negative wave drag from planetary-scale Rossby waves drives an equator-to-pole mass circulation (Andrews et al. 1987; Holton et al. 1995): at a fixed location (Eulerian frame of reference), the negative wave drag is balanced by the Coriolis torque from the poleward flow; following an air parcel (Lagrangian frame of reference), the negative wave drag is balanced by a loss of angular momentum as the parcel moves towards the pole. Mass conservation then demands upwelling in the tropics and downwelling in the extratropics. This vertical motion leads to adiabatic heating or cooling which is balanced, respectively, by radiative cooling or heating. For this reason the mean meridional mass circulation is often called the diabatic circulation. However this term is something of a misnomer since, for the most part, it is the departures from radiative equilibrium that are induced by the meridional circulation, rather than the converse; radiation responds to wave drag much as a spring responds to an applied force (Fels 1985), and it would be more accurate to call the meri-

---

1 Technical terms in italics are defined in a Glossary at the end of the paper.
dional mass circulation the wave-driven circulation (apart from the exceptions noted at the beginning of this paragraph).

Despite the presence of the wave-driven circulation, static stability in this region is determined principally by radiative equilibrium and is strongly stable because of the vertical distribution of ozone heating up to about 50 km altitude, through what is therefore called the stratosphere. A consequence is that vertical motion in the stratosphere is strongly inhibited, and transport timescales are therefore much longer than in the troposphere, extending to years. It follows that stratospheric air is well isolated from the surface layer, and even relatively weak photochemical and oxidation processes can systematically affect chemical distributions since they can act over long timescales, in a process known as chemical aging. In fact this process represents the only sink for many long-lived chemical species in the atmosphere (e.g., N₂O, CH₄, and CFCs), and controls their atmospheric lifetimes.

The contrast in transport timescales between the stratosphere and troposphere is reflected in the spatial distributions of many chemical species, so much so that the tropopause is sometimes defined chemically. Figure 2 shows a latitudinal cross-section of ozone taken from airborne lidar measurements, depicting a strong contrast between the ozone-poor troposphere and the ozone-rich stratosphere. The overall shape of the ozone distribution reflects the shape of the tropopause—elevated in the tropics and sloping down to lower altitudes in the extratropics; note the correspondence with the thermal tropopause depicted in Fig. 1—while the small-scale structure reflects two-way intrusions associated with transient deformations of the tropopause. The dashed curve in Fig. 3 shows the meridional gradient of the zonal-mean value of N₂O from the Canadian Middle Atmosphere Model (CMAM) on the 320 K isentropic surface, which lies within the troposphere in the tropics and within the stratosphere in the extratropics (see Fig. 1).
From this diagnostic one can see a strong gradient, implying a clear jump, in N\textsubscript{2}O values at approximately 50°/C\textsubscript{14}S and 40°/C\textsubscript{14}N, defining the tropopause (in each hemisphere) at this level. The solid curve in Fig. 3 shows instead the meridional gradient of the mode (most likely value, i.e., peak) of the probability distribution function of N\textsubscript{2}O from the Canadian Middle Atmosphere Model on the 320 K isentropic surface during a model January. Northward is defined as positive. The 320 K isentropic surface lies within the troposphere (where N\textsubscript{2}O is relatively high) in the tropics and within the stratosphere (where N\textsubscript{2}O is relatively low) in the extratropics. The narrow regions of large positive and negative gradients correspond to the tropopause. Courtesy of Sunny Arkani-Hamed, University of Toronto.

From the above discussion it is clear that there are several different ways to characterize the stratosphere and troposphere, and their definitions are thus somewhat fuzzy (see Haynes and Shepherd 2001). In the tropics, moist convective adjustment holds only up to about 11–12 km or so, well below the thermal tropopause at about 16–17 km. Palmén and Newton (1969) referred to the former level as the “secondary tropical tropopause”, and identified it with the maximum Hadley-cell outflow. Above this level, the extent of convective penetration decreases rapidly with altitude and there is a continuous transition in the thermodynamic balance between the “tropospheric” tropical regime (convective adjustment, with zero net adiabatic heating/cooling and with radiative cooling in clear-sky descent balancing convective heating) and the “stratospheric” tropical regime described earlier (adiabatic cooling in clear-sky ascent balancing radiative heating). Based on such thermodynamic considerations Thuburn and Craig (2000) argued that the region between the secondary tropical tropopause and the thermal tropopause—sometimes called the “tropical tropopause layer”—is essentially “stratospheric” in character, and that the thermal tropopause has no particular radiative significance. Chemically speaking this region is also transitional, since contact with the boundary layer is not especially recent and air can be chemically aged to some extent. However the thermal tropopause retains a particular chemical significance because of its implications for water vapour through the freeze drying mechanism.

In the extratropics, there is a similar ambiguity. At any given time the tropopause can be deformed, but at some point stratospheric air can become “tropospheric”, or vice-versa, and stratosphere-troposphere exchange irreversible. However the timescale on which this occurs may be different for chemical species than for potential vorticity (which is the most useful dynamical marker of the tropopause); in individual soundings the ozone-defined tropopause is generally located below the dynamical (or thermal) tropopause (Bethan et al. 1996). The seasonal cycle of ozone provides yet another perspective. Stratospheric planetary-wave drag, and hence the equator-to-pole wave-driven circulation in the stratosphere, is largely restricted to the winter season, because only then does a westerly zonal flow permit significant planetary-wave propagation into the deep stratosphere (Charney and Drazin 1961). The transport of chemical species associated with the equator-to-pole circulation in the stratosphere, known as the Brewer-Dobson circulation, transports ozone from its photochemical...
production regions in the tropical upper stratosphere to the extratropical lower stratosphere. This leads to the observed wintertime buildup in ozone in the extratropical lower stratosphere, with maximum ozone values in spring and relaxation towards photochemical control through the summer (Randel et al. 2002). In the tropical upper troposphere, on the other hand, ozone exhibits a maximum in summer, presumably because of local photochemical production (Logan 1999). Intrusions of tropical upper tropospheric air into the lowermost stratosphere modify the chemical composition of the air descending in the latter region, to the extent that in high northern midlatitudes the seasonal cycle of ozone exhibits a summer maximum at and immediately above the thermal tropopause (Logan 1999).

2. The upper troposphere/lower stratosphere region

It is clear from the above discussion that one should probably abandon any notion of the tropopause. In fact, scientific interest focuses more generally on the upper troposphere/lower stratosphere (UT/LS) region. Dynamically, this is the “middleworld” (Hoskins 1991), defined as the region wherein isentropic surfaces do not lie entirely within the stratosphere, nor do they intersect the surface layer (Fig. 1). From a transport point of view this means that timescales for contact with the surface layer are not particularly short, but are not particularly long either—say, weeks to months. The UT/LS region is the physical interface between the deep troposphere and the deep stratosphere (the “overworld”, where isentropic surfaces lie entirely within the stratosphere); stratosphere-troposphere exchange by definition occurs here. Transport through this region is mainly upward in the tropics and mainly downward in the extratropics (the Brewer-Dobson circulation), but there is significant two-way exchange, particularly in the extratropics. As noted earlier, the tropospheric Hadley circulation is largely confined below this region.

The UT/LS region has some particular characteristics that make it distinctive. First, radiative timescales are relatively long, because of a near balance between thermal infrared cooling and heating from the different greenhouse gases (Clough and Iacono 1995). This means that temperature is highly sensitive both to dynamical forcing by adiabatic warming or cooling (for given radiative relaxation) and to radiative changes (for given dynamical forcing), allowing for the possibility of strong climate sensitivity. The radiative sensitivity is particularly delicate because of the key role in this region of the relatively short-lived greenhouse gases ozone and water vapour, which are strongly affected by transport. Second, chemical timescales are relatively long, partly because of the absorption of UV radiation by the stratospheric ozone layer above (which reduces the photochemical forcing) and partly because of the relatively low temperatures (which tend to reduce chemical reaction rates). As with temperature, this means that chemical distributions are highly sensitive both to what one might call “dynamical forcing” by transport (for given chemical rates) and to changes in chemical rates (for given dynamical forcing). The chemical sensitivity is particularly delicate because the low temperatures imply the role of condensed matter (solid and liquid clouds and aerosols) which can drastically alter the chemical partitioning.

There are two broad ways in which the UT/LS region affects climate. The first is through the structure of the tropopause itself. For example, the cold-point temperature at the tropical tropopause (Fig. 1) controls water vapour values in the stratosphere by almost entirely dehydrating the moist tropospheric air as it enters the stratosphere. Although details of the process remain controversial (see section 6), there is no doubt that Brewer’s (1949) freeze drying mechanism is essentially correct. This control over stratospheric water vapour has a direct impact on stratospheric chemistry, especially ozone chemistry. The structure of the tropopause also affects climate dynamically, because the meridional temperature gradients associated with the sloping midlatitude tropopause (Fig. 1) are linked, via thermal-wind balance, with the zonal wind field at the tropopause and above. The properties of the zonal wind field determine the propagation characteristics of atmospheric waves, and thereby influence the propagation of planetary waves into the wintertime stratosphere (thus affecting, for example, polar ozone; see section 7), the trapping of stationary waves within the tropo-
sphere (thus affecting regional climate), and the location of synoptic-scale wave drag (thus affecting surface pressure; see section 8).

The second broad way in which the UT/LS region affects climate is through transport and mixing of chemical species. The role of transport in determining the spatial distribution of the short-lived greenhouse gases ozone and water vapour has already been mentioned. Since ozone also absorbs harmful UV radiation, its abundance affects the amount of such radiation that can reach the Earth’s surface, with implications for the health of the biosphere (including humans). An additional example of climate impacts is that of emissions from commercial aircraft flying in the UT/LS region. The fate of these emissions in terms of stratospheric chemistry very much depends on the transport pathways to the upper stratosphere.

However the entire stratosphere influences the UT/LS region through nonlocal couplings that are dynamical, chemical, and radiative. Dynamically, the mechanism of downward control (Haynes et al. 1991) expresses the effect of stratospheric wave drag at a given altitude on temperatures below through the diabatic circulation, in the absence of time dependence (e.g., at solstice, or in the annual mean). Equivalently, the departure from radiative equilibrium at a given level is (in the absence of time dependence) determined by wave drag above the level of interest. Chemically, the Brewer-Dobson circulation brings chemical species from the middle and upper stratosphere down to the extratropical lower stratosphere. There is an additional nonlocal chemical coupling in the vertical through filtering of the UV radiative fluxes that drive photochemistry, so that chemical changes at one level affect chemical distributions below. (This is why ozone loss in the upper stratosphere leads to more ozone in the lower stratosphere, all else being equal, because it allows more UV radiation to penetrate to lower levels.) Radiatively, the same filtering of short-wave fluxes is relevant, and there is also nonlocal coupling through long-wave fluxes in optically-thick bands.

3. Stratosphere-troposphere coupling

As noted earlier, the global-mean stratosphere is close to being in radiative equilibrium, and there is very little energy transfer from the troposphere to the stratosphere. However, angular momentum transfer by a wide spectrum of atmospheric waves causes local departures from radiative equilibrium, with associated zonal wind changes, reflected in such diverse phenomena as the quasi-biennial oscillation (QBO) in tropical zonal winds and stratospheric sudden warmings during the Arctic winter, as well as the Brewer-Dobson circulation itself (e.g., Shepherd 2000). (The vertical transfer of angular momentum by geostrophically balanced motions such as planetary waves is, curiously, reflected in the meridional heat flux—see e.g., Andrews et al. (1987)—but it is nevertheless the angular momentum transfer that is the key process involved.) In this respect the troposphere has historically been considered the driver of the stratosphere, and this aspect of stratosphere-troposphere coupling has never been questioned. To a first approximation it comes down to mass: the stratosphere represents only about 10–20% of the mass of the troposphere (depending on one’s definition of the tropopause). Nevertheless, as discussed above the stratosphere is by no means passive, and there are a variety of coupling mechanisms through which the stratosphere can affect the UT/LS region and thus tropospheric climate.

The challenge for our scientific understanding is to understand the current coupling between the stratosphere and troposphere, the natural variability of this coupling, and its sensitivities to external changes.

In the stratosphere, radiative and chemical effects are generally stable, or “dissipative” in dynamical-systems language, meaning that in the absence of external changes the equations do not exhibit chaos but instead collapse onto steady-state solutions. (The one notable exception is the middle atmosphere Hadley circulation.) In the absence of dynamical forcing from wave drag, the stratosphere would be close to radiative and photochemical equilibrium, though with a lag because of the seasonal cycle and the finite radiative and chemical relaxation timescales. (There would be a zonal flow in thermal-wind balance with the temperature field, but it would be essentially irrelevant.) In contrast, the dynamical equations exhibit chaotic solutions even under constant external forcing. It is, therefore, dynamical forcing that primarily leads to interannual variability in
the stratosphere. However, this variability is by no means only dynamical: chemical and radiative variability arise from feedback mechanisms. The QBO in stratospheric ozone is a well-known example (Hasebe 1994; Chipperfield et al. 1994). Of course, the stratosphere can also be perturbed chemically or radiatively, for example by a volcanic eruption. In this case the radiative-dynamical-chemical coupling operates in several directions. Figure 4 shows the results of a simulated volcanic eruption in CMAM. The enhanced aerosol in the tropical lower stratosphere leads initially to local heating and enhanced tropical upwelling, but this is a transient response which dies away after several months. On longer timescales, the enhanced aerosol leads to higher temperatures as the increased absorption of radiation is balanced by increased long-wave cooling, leading to a moistening of the tropical water vapour tape recorder. (In the real atmosphere, the warming of the lower stratosphere observed after Pinatubo did not show up in the tropical tropopause temperature, but Pinatubo appears to have been exceptional in this respect; there was a warming of the tropical tropopause following El Chichon, for example (Randel et al. 2000).) This signal takes several years to propagate into the upper stratosphere, because of the slow timescale of the Brewer-Dobson circulation, so the chemical impact of the eruption lasts for several years. This example illustrates the long timescales that are inherent in the stratosphere, and the global nature of the radiative-dynamical-chemical coupling. It is only fairly recently that this has been appreciated, yet it is absolutely crucial to understanding, for example, the response of stratospheric ozone to volcanic eruptions and separating this from long-term trends (see Chapter 7 of WMO 1999).

A mechanism for interannual variability in the stratosphere is what Scott and Haynes (1998) have termed the “low-latitude flywheel”. The balanced response to an imposed wave drag induces a meridional circulation (Eliassen 1951) which, under radiative damping, evolves to the downward-control solution (Haynes et al. 1991). However the timescale for this evolution becomes extremely long in the tropics (Holton et al. 1995), technically going to infinity at the equator. This is because the small value of the Coriolis parameter means that a given zonal wind signal has a much smaller temperature signal than in midlatitudes, and is therefore only weakly damped by radiation (hence the “flywheel”). Scott and Haynes (1998) showed that this mechanism could lead to interannual variability in a stratosphere-only model; the effect was missed by Yoden (1990) because his midlatitude beta-plane channel could not...
include this specifically equatorial mechanism. The effect is illustrated in Fig. 5, from an idealized zonally-symmetric balance model; a switch-on annually repeating forcing in the extratropics is seen to take several years to equilibrate.

To understand and quantify long-term variability in the stratosphere it is essential to look for patterns of feedbacks and couplings that characterize this variability. Otherwise one is reduced to a purely statistical approach, treating variability as noise, which confounds estimates of long-term changes (partly because the variability is large, and partly because it is not truly separable from the long-term changes). Because dynamics is the driver of stratospheric variability through the wave-driven circulation, this suggests trying to isolate the dynamical effects and relate them to concomitant chemical or radiative effects. One example is the “see-saw” between tropical and extratropical temperatures identified by Yulaeva et al. (1994). Provided the static stability is independent of latitude, conservation of mass ensures that the net adiabatic heating over a pressure surface must vanish. Then provided the radiative relaxation rate is also independent of latitude, this implies that the net departure from radiative equilibrium must also vanish. Since the latitude band 30°S–30°N contains half the mass of the atmosphere, the temperature behaviour within this band must then mirror, with a negative sign, the behaviour outside this band, apart from changes in $T_{\text{rad}}$. Put another way, departures from radiative equilibrium temperatures are induced by the diabatic circulation, which since it is mass-conserving acts as a “see-saw” with opposite effects in the tropics and extratropics. This provides a useful diagnostic for separating radiative from dynamical effects, as is discussed in more detail in section 4.

A second example of a pattern of variability is the link between the diabatic and the Brewer-Dobson circulations (Dunkerton 1978). Dynamically-induced downwelling in the extratropics raises temperatures above radiative equilibrium, but also transports ozone from its photochemical source regions in the tropical upper stratosphere. This gives rise to a positive temperature-ozone correlation in the lower stratosphere (Randel and Cobb 1994), which is to a first approximation entirely coincidental. (It would be true for any tracer that had a spatial distribution similar to that of ozone.) Stronger extratropical downwelling is the reason why the northern hemisphere extratropical stratosphere is both warmer and has more ozone than the southern hemisphere. Radiative feedbacks enhance this positive correlation (more ozone leads to more heating), as do chemical feedbacks in polar regions in late-winter/spring (higher temperatures imply less polar ozone loss; this is the opposite of the effect for gas-phase ozone chemistry, as seen in the upper stratosphere where the negative temperature-ozone correlation is well documented, e.g., in the planetary-wave signature in CRISTA 2 measurements in Ward et al. (2000)). Separating the purely dynamical ozone-temperature correlation from the radiative-chemical feedbacks is a major focus of current research on attribution of observed ozone changes, both at midlatitudes and in the Arctic winter/spring season (see Chapter 7 of WMO 1999).
We can see how some of these couplings come together in a simple thought-experiment. Suppose that for some reason (climate change, a volcanic eruption, or low-frequency climate variability) the amount of stratospheric wave drag decreased. This would weaken both the diabatic and the Brewer-Dobson circulation. The weaker diabatic circulation would warm the tropical tropopause (less ascent) and cool the extratropical lower stratosphere (less descent). The weaker Brewer-Dobson circulation would mean less ozone transport to the extratropical lower stratosphere, providing further cooling from the radiative feedback. Lower polar temperatures would imply a stronger polar vortex, hence more of a barrier to ozone transport (as we see today with the Antarctic as compared with the Arctic), and still less polar ozone (from transport); but, also, more PSC-induced ozone loss. The weaker Brewer-Dobson circulation would also mean that stratospheric air would be more chemically aged, thus increasing water vapour (through methane oxidation) and decreasing methane. The warmer tropical tropopause mentioned earlier would further increase stratospheric water vapour (through a reduction in freeze drying). The increased water vapour would lead to more chemical ozone loss from the HOx catalytic cycles (except perhaps in the middle stratosphere), further decreasing stratospheric ozone. In this example, we see that the various radiative-dynamical-chemical feedbacks tend to be positive (but not actually destabilizing). One might speculate whether this sort of behaviour, together with the known chemical forcings, can account for the rather puzzling “hockey-stick” temporal behaviour observed in the stratosphere by the UARS instruments during the 1990s, where there were apparent linear trends in many fields over the first half of the decade, but essentially zero trends over the second half (Randel et al. 2000).

We now consider some particular current issues in stratosphere-troposphere coupling. In each case the existing state of knowledge is summarized, and current questions and uncertainties identified.

4. Maintenance of the tropopause

As noted earlier, the troposphere may be regarded as the region wherein turbulent heat transfer, driven by the dynamical instability of the radiative equilibrium state, takes place. In the tropics, the heat transfer is mainly achieved by convection. In the extratropics, it is mainly achieved by baroclinic instability (Held 1982). On this basis, one should expect no particular relation between the tropical and extratropical tropopauses and, in fact, the extratropical tropopause at about 8–10 km is considerably lower than the tropical tropopause (Fig. 1). The stratospheric diabatic circulation acts to raise and cool the tropopause in the tropics, and to lower and warm it in the extratropics, thus enhancing this contrast. A cartoon of this mechanism is shown in Fig. 6a, and results from controlled experiments with a gen-

![Fig. 6.](image_url)
Several aspects of this picture remain poorly understood. In the tropics, the region of moist convective adjustment does not reach anywhere close to the tropopause, and the tropical upper troposphere (i.e., the tropical tropopause layer) is a transition layer with mixed characteristics. This layer in some sense shields the tropical tropopause from the surface, making the response of the tropical tropopause to surface changes rather complex, and certainly not as simple as depicted in Fig. 6b. Since the thermal radiative balance in this layer is close to being neutral, with an important component from water vapour and ozone, it is highly sensitive to changes. Our ability to predict greenhouse-gas induced changes in the tropical tropopause with any kind of confidence is therefore far from evident.

The annual cycles of 100 hPa temperature (corresponding roughly to the tropopause in the tropics) simulated in middle atmosphere GCMs exhibit a wide variety of behaviours (Fig. 8). The observed cycle, labelled as ERA-15, shows the “see-saw” behaviour between tropics and extratropics referred to earlier, with the lowest tropical temperatures occurring in boreal winter. None of the 13 models depicted get this
quite right. For example, the CMAM gets the correct kind of annual cycle as well as the see-saw behaviour, with very little net bias, implying that the radiative driving at 100 hPa is about right. However the amplitude of the annual cycle is too strong, with evidently too weak a diabatic circulation in boreal summer and fall. In contrast, the GFDL SKYHI model has about the right amplitude of the annual cycle, but a very strong cold bias in the tropics which points to radiative problems. (Weakening the diabatic circulation would merely shift the cold bias to the extratropics.) Other models don’t get the phase of the annual cycle correct, or fail to...
exhibit the see-saw. With such a wide range of simulated tropopause behaviours, it is difficult to have much confidence in model predictions in this region. This situation surely reflects the sensitivities discussed above; furthermore, GCMs hardly even resolve the tropical tropopause layer with their typical vertical grid spacing of 1–2 km.

The structure and maintenance of the extratropical tropopause is not well understood either. A key issue revolves around what determines the static stability of the troposphere, but this amounts to closing the problem of baroclinic adjustment (Held 1982; Held and Schneider 1999)—namely the equilibration of baroclinic instability—which remains very much an unsolved problem. The sloping transition region between the extratropical and tropical tropopauses, corresponding to the subtropical jet, is a region of enhanced meridional gradients of potential vorticity. Idealized numerical experiments have shown how such enhanced gradients can be created naturally by the mixing associated with baroclinic instability (Haynes et al. 2001). However it is also true that the tropospheric Hadley circulation acts to create a region of near-zero potential vorticity within its extent (Held and Hou 1980), thus also forming a sharp potential-vorticity edge (and westerly jet) at the subtropics. It would seem that both processes must be relevant in the real atmosphere.

When one discusses stratospheric wave drag, attention normally focuses on planetary waves. However, although one tends to think of synoptic-scale wave drag being confined to the troposphere, because of Charney-Drazin filtering (Charney and Drazin 1961), it can penetrate evanescently into the stratosphere as is evident, for example, in tracer distributions (e.g., Fig. 5 of Shepherd 2000), which shows a train of 5–6 cat’s eyes at 21 km altitude associated with breaking synoptic-scale Rossby waves, visible in HNO3 observed by CRISTA). In this way synoptic-scale baroclinic waves play a role in determining the structure of the tropopause through stratospheric wave drag, not just through the in-situ mixing in the lowermost stratosphere discussed in the previous paragraph. There is potentially a strong sensitivity here, through a Holton-Tan–like mechanism whereby changes in the subtropical jet could alter the latitudinal location of the synoptic-scale wave drag, and hence the location and strength of the induced diabatic circulation.

5. Mixing, transport, and stratosphere-troposphere exchange

The traditional view of stratospheric transport and mixing (Andrews et al. 1987), in the zonal mean, is advection by the residual (or TEM) circulation together with diffusion. Given that Rossby-wave drag is expected to always yield a negative torque, and hence drive a poleward residual circulation, this appears to account in at least a qualitative way for the Brewer-Dobson circulation. The classical transport theory is based on the assumption of small-amplitude eddies in the form of a mixing-length approximation (so-called K-theory, where K is a diffusivity). In particular, air parcels are presumed to move only a small distance compared to the background inhomogeneities. The reality, however, is that stratospheric eddies are large-amplitude, breaking spectacularly in what is evocatively called the “stratospheric surf zone” (McIntyre and Palmer 1983). Figure 9 shows an example from CMAM, in the middle stratosphere, where particle displacements span many tens of degrees latitude. Moreover surf-zone stirring leads to sharp edges in tracer profiles on the edges of the surf zone (Sparling 2000). Thus the situation is one where eddy length scales are much greater, not much smaller, than the background length scales, and the mixing-length assumption is therefore problematical. An example of how things might go awry is sketched in Fig. 10; here one might expect the Lagrangian vertical velocity to equal zero at quite a different place than the residual (or diabatic) vertical velocity.

Strong inhomogeneities in stratospheric mixing such as the surf zone arise from the forced, fairly coherent nature of stratospheric disturbances. The situation can be described as a generalized form of chaotic advection (Ngan and Shepherd 1999), wherein chaotic particle paths are generated by quasi-regular flows. This leads to distinctive mixing regions and transport barriers. Thus, recent theory has tended to employ an “air-mass” approach, treating the stratosphere as a collection of distinct regions separated by (leaky) barriers
Although this has led to great insight in terms of mixing across certain control surfaces, it is difficult to put the different regions together into a global picture.

It should be understood that a transport barrier does not necessarily imply a barrier to tracer flux. Consider the example of a diffusive regime with a spatially uniform tracer flux $C_0/kw_y$ (as can easily be arranged by a suitable specification of sources and sinks), where $k$ is the diffusivity and $w_y$ is the mean tracer gradient. In that case a sharp tracer edge ($\partial w_y/\partial y$) would just reflect a minimum in $k$, with no implications in terms of the flux. The sense in which a region of minimum diffusivity would nevertheless represent a transport barrier is that air parcels approaching the region would be more likely to turn around than to move through it (as at the edge of the surf zone). This illustrates the difference between the Lagrangian (parcel) problem and the tracer problem. If, for example, a narrow transport barrier was located between two mixing regions, then

Fig. 9. 30-day particle advection calculation in the southern hemisphere using isentropic winds from the Canadian Middle Atmosphere Model during a model July. The isentropic level is 1000 K which corresponds roughly to 35 km altitude, the middle stratosphere. The particles are initially aligned along latitude circles. Taken from Ngan and Shepherd (1999).

Fig. 10. Schematic of simple example violating the mixing-length hypothesis. The left panel shows the meridional profile of the diabatic vertical velocity $\bar{w}_d$ (solid curve) together with a transport barrier (dashed line). The right panel shows the meridional profile of the resulting Lagrangian vertical velocity $\bar{w}_L$, assuming that the mixing is much more rapid than the vertical motion. There is a region where the signs of the two vertical velocities are opposite. The discrepancy arises because the mixing length is much greater than the length scale over which the mean properties vary.
the Lagrangian velocity (defined in terms of parcel origin) would diverge at the barrier: air parcels located just on either side of the barrier would tend to move off, in opposite directions, into the adjacent mixing regions. On the other hand the tracer flux through this region would depend solely on the distribution of sources and sinks, and could well be spatially uniform.

In fact, it is not clear whether the traditional view of transport and mixing is all that bad. Comparisons of the Lagrangian and residual circulations in a mechanistic simulation of a planetary-wave surf zone show that the two vertical velocities are actually surprisingly similar (Pendlebury 2001). The two meridional velocities agree closely where the waves are not breaking, as one might expect since the concept of Stokes drift is relevant in such regions. On the other hand it is much harder to reconcile the meridional velocities in the surf zone, because of the sensitivity of the Lagrangian meridional velocity to the way in which one labels air parcels. (An example of the issues involved was noted in the previous paragraph.) Thus there is still a question as to how to quantify the Brewer-Dobson circulation.

An important (and closely related) issue is how to quantify stratosphere-troposphere exchange. It is clear that a very significant fraction of the exchange across the extratropical tropopause is two-way. There is a net poleward mass flux in the UT/LS region (in isentropic coordinates (Held and Schneider 1999) or in the TEM formulation (Edmon et al. 1980)), but the ozone flux, for example, is equatorward and therefore presumably must be accomplished by the two-way exchange. On the other hand, there is no such discrepancy when one considers vertical fluxes across the 380 K isentropic surface at the top of the lowermost stratosphere (Fig. 1): both mass and ozone are descending monotonically. If one is interested in long-lived chemical species, then the bulk downwelling is probably sufficient for a determination of stratosphere-troposphere exchange. However, if one is interested in faster chemical processes, as is likely the case with upper tropospheric ozone, then the details of the two-way exchange could matter. The value of the classical transport theory for quantifying the ozone flux under such circumstances remains to be ascertained.

6. Tropical upwelling and dehydration

The Brewer-Dobson circulation involves persistent tropical upwelling throughout the year, with maximum upwelling in boreal winter and minimum upwelling in austral winter. This upwelling cools the tropical tropopause, more so in boreal winter than in austral winter, leading to the annual cycle in tropical tropopause temperature seen in the ERA-15 data in Fig. 8. This annual cycle modulates Brewer’s (1949) freeze drying mechanism, which together with the persistent upwelling leads to the celebrated tropical tape recorder in water vapour of Mote et al. (1996) (and seen in the upper panel of Fig. 4). This much, at least, is known. However, the cause of this annual cycle in upwelling remains a question of debate. Yulaeva et al. (1994) and Holton et al. (1995) argued that it reflects the stronger extratropical planetary-wave drag (the so-called “extratropical pump”) in the northern hemisphere in its winter than in the southern hemisphere in its winter (reflected in the asymmetry in ozone distribution and in polar jet strength, for example). However, downward control suggests that the influence of extratropical wave drag would be limited to the extratropics on seasonal timescales (Plumb and Eluszkiewicz 1999; Semeniuk and Shepherd 2001b). Moreover, if extratropical wave drag does control tropical upwelling, then why is the upwelling observed to be centred on the summer side of the equator? In the upper stratosphere it is possible to account for this latitudinal structure in the upwelling profile through a combination of the solstitial middle atmosphere Hadley circulation (Semeniuk and Shepherd 2001b) and the response to extratropical wave drag (Tung and Kinnersley 2001), but these explanations do not appear to be relevant to the tropical tropopause. It may well be that subtropical synoptic-scale wave drag could play a role, as it does penetrate above 100 hPa (as noted in section 4) and could conceivably reach sufficiently close to the equator to produce significant tropical upwelling. Of course, one would then need to find a mechanism for the interhemispheric asymmetry.

Quite apart from the large-scale issues, there remain questions about what sets the entry-level water vapour of the stratosphere. Although Brewer’s mechanism is clearly roughly
correct, making quantitative progress has proven to be remarkably difficult. In particular, it does not seem possible to relate the observed water vapour to the observed temperatures in a simple manner. There continue to be debates about whether the “final” dehydration (setting the entry-level value of stratospheric water vapour) is achieved through convective overshoots or through large-scale ascent (Sherwood and Dessler 2000). Even if it is the latter, because of longitudinal asymmetries in tropical tropopause temperatures it is important to know transport pathways into the stratosphere. It is not known, for instance, whether one can safely assume that all air parcels pass through the region of lowest temperatures. Unfortunately the poor quality of analyzed winds in the tropics makes such calculations difficult. Moreover, the microphysics of dehydration are uncertain. Even if water vapour condenses into ice, unless the ice falls out—and this is far from obvious—the air will simply be rehydrated once it reaches the warmer stratosphere. There are likewise questions about whether the nucleation is homogeneous or heterogeneous, which affects the threshold temperatures. Until these issues get resolved, our quantitative understanding of the response of stratospheric water vapour to changes in tropical tropopause temperatures will remain limited.

A major current puzzle is the apparent increase of water vapour in the stratosphere since the 1960s (Rosenlof et al. 2001). Tropical tropopause temperatures have if anything been decreasing during this period (Seidel et al. 2001), which would be expected to reduce stratospheric water vapour, subject to the uncertainties discussed above. Both the observed increase in methane, and a hypothetical slower Brewer-Dobson circulation, could in principle increase stratospheric water vapour, but these mechanisms do not appear to be adequate (Rosenlof 2002).

7. Stratospheric wave drag

Stratospheric wave drag drives the Brewer-Dobson circulation, and warms and weakens the wintertime polar vortices. Most of this drag, in the extratropics, is believed to come from planetary waves. The strength and location of planetary-wave drag can easily change, both interannually (e.g., through the Holton-Tan mechanism) and on very short timescales (e.g., in stratospheric sudden warmings, in the Arctic). The effect of gravity-wave drag is known to be important in the Antarctic; all middle atmosphere GCMs without some kind of artificial damping are prone to a very significant cold bias in the Antarctic vortex, unless they include a parameterization of (non-orographic) gravity-wave drag that is able to generate sufficient downwelling and adiabatic warming over the pole. The importance of gravity-wave drag in the Antarctic, even for temperatures in the lower stratosphere, is understandable in light of the relative absence of planetary-wave drag at high latitudes; even relatively small amounts of gravity-wave-induced downwelling can create significant departures from radiative equilibrium, because of the long radiative damping times in the polar night (Garcia and Boville 1994).

The conventional wisdom has been that gravity-wave drag is less important in the Arctic, simply because it is overwhelmed by planetary-wave drag. However, some results have suggested that gravity-wave drag is important in the Arctic too. Boville (1995) found that when the NCAR MACCM was run without any gravity-wave drag parameterization, it developed a cold-pole problem in the Arctic winter. Beagley et al. (1997) diagnosed the contribution to lower stratospheric downwelling over the Arctic polar cap in simulations with the CMAM, and found that the orographic gravity-wave drag parameterization accounted for nearly one-half of the downwelling. Given the uncertainties associated with gravity-wave drag parameterization, most of which involve a proper understanding of the sources, this sensitivity introduces concerns about how well we can model both the Antarctic and the Arctic stratospheres.

Uncertainties in modelling the Arctic stratosphere are especially problematical because the Arctic exhibits a high degree of variability, as manifested most spectacularly in stratospheric sudden warmings. To get reliable statistics in middle atmosphere GCM simulations, even with no interannual variations in external forcing, appears to require simulations of at least several decades. Because this timescale overlaps with that of anthropogenic forcing, the problem of predicting the transient response to
such forcing is a delicate one indeed; single transient simulations are clearly of no value under such conditions. (On the other hand, the observed record consists of only a single realization, making the attribution issue thorny as well.) For example, the observed decrease in Arctic late-winter/spring ozone through much of the 1990s was initiated by changes in Arctic wintertime conditions, with the Arctic vortex being particularly strong and cold during this period (see Chapter 7 of WMO 1999). That these changes in the Arctic vortex could not be the result of the ozone decreases themselves is clear both from the timing of the changes (WMO 1999) and from GCM experiments with the imposed ozone decreases (Langematz 2000; Rosier and Shine 2000). They also could not be the direct (radiative) result of greenhouse gas changes, which are far too weak. Rather, the stronger and colder Arctic vortices arose from reduced wave drag, leading to reduced polar ozone through both dynamics (less transport into the vortex) and chemistry (more PSCs). (There is a particular sensitivity in polar regions, because of a dynamical feedback whereby a stronger polar vortex tends to shield the pole from planetary-wave drag. It is this feedback, acting in the opposite direction, that leads to stratospheric sudden warmings.) The question is then why the Arctic conditions changed. There were claims that this was a dynamical result of greenhouse-gas induced global change (Shindell et al. 1998; Salawitch 1998). However, 1998 and several winters since have been much warmer, suggesting that this behaviour was more likely the result of low-frequency variability.

Similar issues arise in midlatitudes. Variability in planetary-wave drag is reflected in variability in ozone transport to midlatitudes (Randel et al. 2002), although probably with rather less sensitivity than in polar regions. It remains controversial what fraction of the observed decreases in midlatitude ozone over the last 20 years might be attributable to changes in circulation, and what fraction of these latter changes are secular as opposed to natural variability. Until we obtain a better understanding of stratospheric variability it will be difficult to address these questions in a quantitative fashion. Looking at patterns of variability and couplings between different fields, as discussed in section 3, will probably be crucial in this respect.

8. Annular modes of variability

There is observational evidence of coupled high-latitude variability between the troposphere and stratosphere, mainly in the zonally-symmetric component of the flow (Thompson and Wallace 1998). This phenomenon is known either as the Arctic and Antarctic Oscillations (the AO and AAO) or as the northern and southern hemisphere annular modes, although it should be cautioned that there is no oscillation or bimodality and the opposite phases of the modes simply correspond to the wings of the relevant PDF. The picture in the northern hemisphere is complicated by the presence of stationary waves, but it is believed that the coupling mechanisms between the troposphere and stratosphere are the same in the two hemispheres. There appears to be evidence for downward propagation in the signal (Baldwin and Dunkerton 1999), raising the question of whether the stratosphere is directly influencing the troposphere through this mode of variability. Given the similarity between the AO and the North Atlantic Oscillation (NAO), and the demonstrated role of the latter in tropospheric climate variability (Hurrell 1995), the AO and AAO have potentially important implications for the understanding of tropospheric climate variability.

A basic dynamical constraint is that waves generally propagate upward, while zonal-mean anomalies generally propagate downward. The latter occurs in several ways. First, the instantaneous balanced response to wave drag involves both a change in the zonal wind and an induced meridional circulation (Eliassen 1951). The induced meridional circulation involves a change in the mass distribution, and is reflected in a change in the surface pressure (Haynes and Shepherd 1989). Another way of looking at this is that the response to the applied torque from wave drag goes partly into changes in relative angular momentum (through the zonal flow), and partly into changes in planetary angular momentum (through the mass distribution, reflected in surface pressure). Thus, there is an immediate connection between stratospheric wave drag and surface pressure. Second, if the wave drag
is maintained then the induced meridional circulation burrows downward on radiative timescales according to downward control (Haynes et al. 1991). Third, there can be evolution on other timescales by wave—mean-flow interaction, as occurs for example with the QBO. It is important to note that in all cases, any downward propagation is using wave forcing that ultimately came from below, so it is not a case of the stratosphere driving the troposphere; there is no conflict with the notion that the troposphere is too massive to be directly forced by the stratosphere.

Hitherto, analyses of the AO and AAO have used statistical measures such as EOF indices. Since the phenomenon is clearly dynamical, involving wave drag and wave—mean-flow interaction, it is important to try to quantify it with diagnostic tools based on the equations of motion—meaning, in this case, angular momentum and angular pseudomomentum (or Eliassen-Palm wave activity). In this way the actual mechanisms driving the AO and AAO could be investigated, and various hypotheses tested. In particular, angular pseudomomentum should allow one to identify the upward propagation of waves (which would in principle be invisible in the AO and AAO indices, although there could be some component projecting onto the AO because of stationary waves) prior to the downward propagation of zonal-mean anomalies. This kind of analysis might help to explain why this phenomenon is apparently confined to high latitudes; does the polar night jet act as a waveguide, much as the equator acts as a waveguide in the case of the QBO?

The observations show instances where a large surface response occurs nearly instantaneously once the AO or AAO signal reaches the tropopause. This raises the question of whether the stratosphere might somehow condition the troposphere. A possible mechanism involves synoptic-scale wave drag. Thorncroft et al. (1993) noted the strong sensitivity of the location of upper tropospheric synoptic-scale wave drag in idealized simulations of baroclinic life cycles: depending on the strength of the zonal flow, the wave drag occurred either equatorward (and in the upper troposphere) or poleward (and in the middle troposphere) of the subtropical jet. Because of the avoidance of the jet itself, there is a bifurcation between the two possibilities, providing enormous sensitivity. Observations clearly reveal the existence of both kinds of behaviour (Fig. 11); the location of upper tropospheric synoptic-scale wave drag is evidently bimodal. Thus one might imagine that even with a fixed source of wave activity, once a change in the zonal flow reached the tropopause it could alter the subtropical jet, and flip the synoptic-scale wave drag from subpolar to subtropical latitudes (or vice-versa). This would have an immediate and significant impact in terms of surface pressure (Haynes and Shepherd 1989).

There has been much interest in shifts in the AO and AAO indices in the observational record, and in possible shifts under climate change. A possibly provocative question is whether such shifts are anything more than a complicated way of looking at a change in the mean climate.

9. Conclusion

The stratosphere and troposphere are coupled in many ways. Interannual variability arises mainly from dynamics, but is often coupled to radiative and chemical effects. For example, the chemical memory of a volcanic eruption is several years, since the effects are transported through the Brewer-Dobson circulation. The stratosphere is forced by wave drag from tropospheric waves, but the strength and location of this wave drag are highly variable, on all timescales. Given the amplitude of the variability, it is essential to understand cause-and-effect patterns of variability, to separate natural from forced changes. One cannot simply rely on statistical approaches. We also need to understand how these mechanisms operate in middle atmosphere GCMs, in order to develop confidence in our ability to attribute past changes and to predict future changes.

Acknowledgements

The author is grateful to the Kyoto University Foundation for supporting his attendance at what was a most enjoyable workshop. His research is supported by the Natural Sciences and Engineering Research Council, the Meteorological Service of Canada, and the Canadian Foundation for Climate and Atmospheric Sciences. Helpful comments from Sylvie Gravel, Jim Holton, Kirstin Krüger, Paul Kushner, Mi-
Appendix

Glossary

Brewer-Dobson circulation: The chemical transport circulation of the stratosphere. Brewer (1949) inferred a tropical entry to the stratosphere based on the water vapour distribution (see freeze drying mechanism), while Dobson (1956) inferred poleward transport within the stratosphere based on the ozone distribution. Both characteristics match the diabatic and residual circulations, so there is a tendency to use all three terms synonymously. However, chemical transport involves both the mean mass circulation and two-way mixing (i.e., without net transport of mass), and only the former is related to the diabatic or residual circulation. Thus, for example, air in the extratropical lower stratosphere is not as chemically aged as it would be if it were advected purely by the diabatic or residual circulation (Hall and Plumb 1994).

Chemical aging: The persistent conversion of one chemical species to another by photochemistry (chemistry associated with the absorption of UV and visible radiation) or oxidation. One might reasonably include radioactive decay as well. Air that enters the stratosphere through the tropical tropopause tends to remain in the stratosphere for a considerable time. Because all air in the deep stratosphere (above the lowermost stratosphere) has to have passed through the tropical tropopause, one may define the “age” of stratospheric air, at a given location, as the time since it passed through the tropical tropopause. In fact there is a PDF of ages, because of mixing (Hall and Plumb 1994); a given air parcel is composed of molecules with different ages. The mean of this PDF is the mean age (or simply age, for short). In the upper stratosphere, the age of air is several years. Photochemical and oxidation processes can therefore act on stratospheric air...
over long timescales, and the age of air is inversely related to the concentrations of long-lived species whose primary loss process occurs in the middle atmosphere. In fact, the spatial distributions of such species and of age tend to coincide.

**Diabatic circulation:** The mean meridional circulation inferred from the steady-state thermodynamic equation in isentropic coordinates, where eddy fluxes and meridional advection of heat vanish by construction (Andrews et al. 1987). There is then a balance between adiabatic heating or cooling from vertical flow and diabatic cooling or heating, which in the stratosphere is mainly from radiative processes. The associated meridional flow is inferred from mass conservation.

**Downward control:** The constraint between wave drag and the residual (or TEM) circulation under steady-state or time-mean conditions (Haynes et al. 1991). Under these conditions, the TEM formulation of the zonal momentum (or angular momentum) equation represents a balance between the applied torque from wave drag or frictional drag and the Coriolis torque from meridional flow. The associated vertical flow is inferred from mass conservation. The key assumptions of downward control are that radiative damping can accommodate any steady vertical flow, and that frictional drag is restricted to the planetary boundary layer where it can accommodate whatever steady meridional flow is demanded by mass conservation. It then follows that the vertical flow at any given altitude and latitude is determined by the wave-drag distribution above that location. A corollary of downward control is that a radiative heating anomaly cannot lead to a steady meridional circulation.

**Freeze-drying mechanism:** The process by which air is dehydrated as it enters the stratosphere in the tropics. Because the tropical tropopause is a local minimum of temperature, it also represents a minimum in the saturation mixing ratio of water vapour over water or ice (it turns out that ice is the relevant factor here). Thus, the water vapour in a parcel ascending through the tropical tropopause region will continuously condense into ice clouds. Provided the ice is somehow removed, the dehydrated air will then enter the warmer stratosphere and remain undersaturated. Water vapour is produced within the stratosphere by methane oxidation, but the quantity \( \text{H}_2\text{O} + 2\text{CH}_4 \) (which is conserved under methane oxidation) is remarkably uniform throughout the stratosphere. The only exception occurs in the Antarctic lower stratosphere in winter, and in unusually cold winters in the Arctic, where the extremely low temperatures cause further condensation and dehydration through water-ice PSCs.

**Holton-Tan mechanism:** A mechanism by which the quasi-biennial oscillation (QBO) in tropical zonal winds modulates the strength of the wintertime polar vortex. Planetary Rossby waves tend to break in critical layers, where the phase speed of the wave matches the zonal flow speed (McIntyre and Palmer 1983; see also Shepherd 2000). Since planetary waves tend to be quasi-stationary (because their principal forcing mechanisms are stationary), their critical layers correspond to zero-wind lines. During wintertime, when the extratropical zonal winds are westerly and planetary waves are able to propagate into the stratosphere, the latitude of the zero-wind line depends on the phase of the QBO: in the easterly phase it lies within the winter hemisphere while in the westerly phase it lies within the summer hemisphere. Holton and Tan (1980) realized that this would affect the latitude of planetary-wave drag, and hence the extent of wave-induced downwelling over the pole. In the easterly phase of the QBO, polar downwelling is greater and the vortex weaker and warmer than in the westerly phase. One can generalize this concept and speak of a “Holton-Tan–like mechanism” whereby changes in zonal winds lead to changes in the latitude of Rossby-wave critical layers, and hence to changes in the location of Rossby-wave drag—whether synoptic-scale or planetary-scale.

**Middle atmosphere Hadley circulation:** A thermally forced solstitial circulation in the middle atmosphere. The tropospheric Hadley circulation is driven by convective heating. In the middle atmosphere (stratosphere and mesosphere), in contrast, radiation acts to relax the temperature back to radiative equilibrium \( T_{\text{rad}} \); in the absence of wave drag one would expect \( T \approx T_{\text{rad}} \), in gradient-wind balance with an associated zonal flow \( u_{\text{rad}} \), and no meridional circulation apart from that generated by time
dependence in $T_{\text{rad}}$ (Garcia 1987). However, the gradient-wind balance relationship is nonlinear in $u$, and not every $T$ distribution has a solution for $u$; in particular, the meridional temperature gradient at the equator is required to vanish. At solstice seasons, $T_{\text{rad}}$ has a non-zero meridional gradient at the equator, implying that radiative equilibrium cannot be realized in this region. A meridional circulation within the tropics develops in response, which has been dubbed the “middle atmosphere Hadley circulation” (Dunkerton 1989). It is the only significant thermally forced steady circulation in the middle atmosphere, and is confined to the tropics during solstice seasons.

**PDF (probability distribution function):** The function describing the frequency distribution of the possible values of a given quantity, normalized such that its integral is unity. For a geophysical quantity (e.g., concentration of a chemical species), one may construct the PDF at a given altitude and latitude by binning all the available measurements at different longitudes and times and constructing a histogram of the concentrations: the value of the PDF at each concentration value is the likelihood of observing that concentration value in a random measurement (at the given altitude and latitude).

**PSCs (polar stratospheric clouds):** Liquid or solid aerosols that form at sufficiently low temperatures (achieved only in polar winter). Heterogeneous or multiphase reactions in or on these aerosols convert halogen species (mainly chlorine and bromine) from inactive to active forms, which are capable of destroying ozone through catalytic cycles. PSCs (together with enhanced halogen species) are the essential ingredient in the chemical ozone loss that creates the Antarctic ozone hole every austral spring.

**Residual (or TEM) circulation:** The mean meridional circulation in the Transformed Eulerian Mean (TEM) formulation of the governing equations, denoted $\langle \mathbf{v}^* \rangle$. The TEM formulation involves a transformation of the circulation variables that, to good approximation, collects all the eddy forcing terms in the zonal momentum equation, leaving only diabatic heating and mean advection in the thermodynamic equation (Andrews et al. 1987). In that sense it can be considered a (log-)pressure coordinate approximation to the isentropic formulation of the equations. The residual circulation is not quite equal to the (transient) diabatic circulation, but the differences between them are not quantitatively significant. The great merit of the TEM form of the equations is that the eddy forcing term in the zonal momentum equation is the divergence of the flux of a conserved measure of wave activity, the “Eliassen-Palm (EP) flux”, written $\nabla \cdot \mathbf{F}$. It turns out that the EP flux divergence represents the convergence of angular momentum flux, when properly formulated in terms of angular pseudomomentum (Andrews et al. 1987; see also wave drag below). Thus the TEM equations concisely express the physical relation between angular momentum transfer by waves and the induced meridional circulation. In contrast, non-breaking waves cannot transfer heat across isentropic surfaces, and so it is appropriate that the wave forcing terms disappear from the TEM thermodynamic equation.

**Stokes drift:** The Lagrangian motion of particles induced by a time-periodic flow with zero Eulerian mean. The classic example is for two-dimensional water waves, where particles undergo circular motion in the vertical plane, and the forward motion close to the surface (where the wave amplitude is largest) is not cancelled by the backward motion at depth, despite the fact that at a fixed depth the time-mean (Eulerian) flow is zero. A key feature is that the flow is periodic in time, so the concept is most relevant to linear, non-breaking waves. See Andrews et al. (1987) for more discussion.

**Stratospheric sudden warming:** Rapid warming of the wintertime polar stratosphere (over a few days) induced by the polar downwelling (and adiabatic warming) resulting from a focussing of planetary-wave drag over the pole. The phenomenon can also be understood in terms of a breakdown of the polar vortex through the absorption of negative angular momentum from the anomalous planetary-wave drag. Sudden warmings appear to be essentially a chaotic process, and are the primary cause of Arctic wintertime variability. In our present climate, planetary-wave drag is too weak to produce sudden warmings in the Antarctic.

**Tape recorder:** The propagation of tropical tropopause water vapour values into the stratosphere. Persistent upwelling through the
tropical tropopause, combined with the freeze-drying mechanism, implies that the saturation mixing ratio of water vapour at the tropical tropopause gets carried upward into the stratosphere. It follows that changes in the tropical tropopause temperature should change the amount of water vapour entering the stratosphere, with the signal propagating upward, much as a tape recorder marks a tape; the vertical profile of water vapour in the tropics then provides a record of the time history of the saturation mixing ratio at the tropopause. This effect is demonstrated most clearly in the annual cycle, where the annual cycle of tropical tropopause temperatures gets imprinted on the ascending water vapour signal (Mote et al. 1996). In fact, water vapour is produced within the stratosphere by methane oxidation, so the tape recorder signal is best seen in H$_2$O + 2CH$_4$.

**Wave drag:** Angular momentum transfer by atmospheric waves, whether gravity waves, Rossby waves, or equatorial waves. It is essential to realize that angular momentum transfer by waves is not equal to the flux of wave angular momentum. Rather, one has to begin from the conservation law for angular momentum and derive what is known as the “angular pseudomomentum”, representing the contribution by the waves (see e.g., Shepherd 1990). In the atmosphere, wave drag is represented by the Eliassen-Palm flux divergence (see *residual circulation*), which is a combination of wave momentum and heat fluxes, even though it is angular momentum and not heat that is being transferred by the waves! (The heat-flux contribution to the EP flux occurs through the deformation of isentropic surfaces, much as topography deforms a flow, in a process known as “form drag”.) The sign of wave drag is related to the phase speed of the waves relative to the mean flow, so planetary waves always give a negative drag, equatorial Kelvin waves always give a positive drag, and gravity waves can give a drag of either sign. Wave drag tends to act in a decelerative sense (hence the name), but not always, as for example with the quasi-biennial oscillation.

**Wave-driven circulation:** The meridional circulation induced by the applied torque from wave drag. The instantaneous response is described by Eliassen’s (1951) theory of a balanced response to a mechanical force: the applied torque cannot go entirely into zonal-wind acceleration, which would violate thermal-wind balance, so it also induces a meridional circulation with associated wind and temperature tendencies to maintain a balanced response. In the steady-state limit, the response evolves into the *downward control* state. The latter is a special case of the "gyroscopic pumping" phenomenon (McIntyre 2000), whereby a persistent force applied to a rapidly rotating fluid produces mass transport in a direction perpendicular to the force. (Ekman pumping is another example.) The wave-driven circulation is that part of the *residual circulation* driven by wave drag, although this definition is not without ambiguity except in special cases (Semeniuk and Shepherd 2002).

### References


Dobson, G.M.B., 1956: Origin and distribution of the

Dunkerton, T.J., 1978: On the mean meridional

Eliassen, A., 1951: Slow thermally or frictionally

Edmon, H.J. Jr., B.J. Hoskins and M.E. McIntyre,


——— and ———, 2001: Report on the SPARC Tro-


