The descent of tropospheric air into the stratosphere

By M. N. JUCKES*

University of Munich, Germany

(Received 5 June 1998; revised 29 January 1999)

SUMMARY

Since the stratosphere lies above the troposphere, naive expectation would associate mass transfer from the latter into the former with ascent. This note shows that, in the context of the mass circulation associated with large scales in mid latitudes, the transfer into the stratosphere can actually be associated with descent relative to isentropes. The diabatic effects which accompany the transfer of air from the troposphere to the stratosphere are such as to give the air a downward motion, but, the motion of the tropopause is faster so that the descending air is overtaken by the tropopause and transferred into the stratosphere. Combined with the conclusions of an earlier study, these results suggest that the dryness of a substantial portion of the air in the lowermost extra-tropical stratosphere is caused by its passage through extremely low tropopause temperatures associated with transient mid-latitude high-pressure systems.

KEYWORDS: Atmospheric dynamics Mid latitudes Tropopause

1. INTRODUCTION

This study continues the work of Juckes (1997), hereafter J97, in trying to establish a theoretical framework to describe the long-term mean mass fluxes through and around the extra-tropical tropopause and to quantify the relationship between those mass fluxes and the meridional heat fluxes of large-scale eddies. The approach is based on conservation of mass, momentum and potential vorticity. Using these globally conserved quantities it is possible to establish diagnostic results that do not depend on the local details of the complex dynamics of the tropopause. J97 used quasi-geostrophic theory, as extended to account for vertical displacements of the tropopause by Juckes (1994). This paper generalizes some results of the primitive equations and extends the analysis of the diabatic vertical motion near the tropopause.

A key result of J97 was the link between the mean poleward mass flux in the upper troposphere and the mass flux through the extra-tropical tropopause. When viewed in isentropic coordinates the poleward heat flux in the mid-latitude troposphere consists of net equatorward motion on isentropes in the lower troposphere, and poleward motion on isentropes in the upper troposphere (Johnson 1988; Iwasaki 1989; Juckes et al. 1994; Bartels et al. 1998). Isentropes in the upper troposphere generally intersect the tropopause at some point around a latitude circle, so the mean poleward mass flux on these isentropes is an average over both tropospheric and stratospheric air. J97 made a theoretical analysis of the poleward mass flux and deduced an equatorwards mass flux in the stratospheric portion of these isentropes. This result is somewhat surprising, since it is generally accepted that air enters the stratosphere in the tropics, then flows polewards in the stratosphere before sinking back into the troposphere at higher latitudes. The latter circulation pattern was proposed by Brewer (1949), who acknowledged that the angular momentum budget was problematic. Brewer’s original hypothesis suggested a circulation extending no more than a few kilometres above the tropopause, but subsequent literature has dealt mainly with the stratosphere above 100 hPa. The region between 100 hPa and the tropopause has been studied less because its vicinity to the troposphere makes analysis difficult, be it observational, numerical or theoretical.

The nature of the circulation above 100 hPa has been substantially clarified with the help of the transformed Eulerian-mean theory of Andrews and McIntyre (1976).

* Corresponding author, present address: Atmospheric, Oceanic and Planetary Physics, Clarendon Laboratory, Parks Road, Oxford OX1 3PU, UK.
Combining this theory with observations (see section 3 for literature) implies a mass flux of $8 \text{Tg s}^{-1}$ ($1 \text{Tg s}^{-1} = 10^9 \text{kg s}^{-1} = 3.15 \times 10^{16} \text{kg yr}^{-1}$) flowing towards the winter pole from the tropical stratosphere. This should be compared with the $30 \text{Tg s}^{-1}$ estimated by Brewer. Although the latter was a rough estimate, the factor of four discrepancy requires an explanation. This paper offers one.

The conclusion drawn in J97, that there is an equatorward mass flux in the lowermost stratosphere, follows from two robust properties of the winter circulation. Firstly, warm tropospheric air is associated with a high tropopause while cold tropospheric air is associated with a low tropopause. Secondly, there is a net polewards flux of warm tropospheric air balanced by an equatorwards flux of cold tropospheric air. These two facts taken together imply a net equatorwards flux of stratospheric air. In J97 this flux was quantified in the context of quasi-geostrophic tropopause (QGT) theory (Juckes 1994). Below, part of that derivation is generalized to the primitive equations.

Given an equatorward mass flux in the lowermost stratosphere, closure of the mass budget then requires that there be a transfer of air from the troposphere into the stratosphere at high latitudes, as found by Hoerling et al. (1993) and Grewe and Dameris (1996). The analysis of J97 shows that this transfer is consistent with the exchange rates expected during the non-conservative decay of baroclinic eddies.

The analysis in J97 is largely in the framework of QGT theory, which provides a quantitative link between potential-temperature anomalies on the tropopause, created by meridional advection, and the anomalies in tropopause height. The use of QGT theory to analyse this circulation is justified by the low Rossby number of the synoptic-scale eddies which drive it and by the fact that the meridional mass flux in isentropic coordinates is dominated by the geostrophic component given by the correlation between the geostrophic meridional velocity and the isentropic-coordinate pseudo-density (Johnson 1988; Juckes et al. 1994). The analysis of J97 makes use of the results of Juckes (1994) showing how quasi-geostrophic theory can be applied/extended to the dynamics of the tropopause. The standard version of quasi-geostrophic theory does not cover this situation because of the large horizontal variations in static stability which are created through vertical displacements of the tropopause. By explicitly accounting for the vertical motion of the tropopause this problem is circumvented. Juckes (1999) uses an analysis of shear-line structure to evaluate the accuracy of QGT theory over a range of Rossby numbers and finds quantitative accuracy up to about Rossby number 0.3.

The zonally averaged circulation predicted by this analysis enters the stratosphere at high latitudes, flows equatorwards in the lowest layers of the stratosphere and then returns to the upper troposphere where it flows polewards. The zonal mean is a useful diagnostic because it is well defined and readily evaluated, but its interpretation is not always clear. J97 also considered a quasi-Lagrangian mean defined as an average around contours of potential temperature on the tropopause. The variation of air-parcel characteristics over the averaging volume is then small, so the average should also be typical of individual air parcels within the volume averaged. The zonal mean, on the other hand, averages over air parcels with widely varying properties and so is not necessarily typical of any. The theory in J97 predicts that this quasi-Lagrangian-mean mass flux is directed from the stratosphere to the troposphere at high latitudes. The difference between the zonal mean and the quasi-Lagrangian mean highlights the importance of meridional displacements in the exchange process.

This paper considers a further subtlety of the upper-tropospheric and lower-stratospheric circulation: namely, that it is in fact thermally direct, even though a mass flux from the troposphere to the stratosphere at high latitudes and vice versa at low
latitudes appears to imply a thermally indirect motion. This implication follows if it is assumed that transfer from the stratosphere to the troposphere is associated with downward motion relative to isentropes, an assumption which, surprisingly, turns out to be unnecessary in the context of the mid-latitude winter tropopause. It will be shown later that this transfer, from stratosphere to troposphere, can take place in association with rising motion, but in such a way that the tropopause rises faster than the air particles, so that there is a net transfer into the troposphere.

Holton et al. (1995) suggest that cross-tropopause exchange is associated with mixing along isentropes. To understand the distinction between that idea and the concepts being presented here it is important to appreciate the distinction between mixing and transport (Green 1970). The following analysis is concerned with the meridional transport of air and its implications for the exchange processes between stratosphere and troposphere. When, for instance, air is transported from low latitudes to a region polewards of the jet it typically forms an anticyclone which is far from being mixed with the environment. Some processes, such as second-order chemical reactions, depend on mixing. For such processes the meridional transport of air does not have a direct influence, though it may be an important precursor. For other processes, however, the transport itself is of importance (the heat and water-vapour budgets, for example). The analysis of J97, continued here, is based on the idea that large-scale eddies dominate the meridional transport in the troposphere.

2. Overview

(a) Five aspects of the meridional circulation

The schematic in Fig. 1 shows five quantitative measures of the mean circulation which feature in the following discussion and indicates how they are related to each other. The quantitative measures are as follows:

(a) The meridional eddy heat flux. From a climatological point of view this could be regarded as the primary characteristic of mid-latitude eddies: the rate at which they transfer heat from the tropics to higher latitudes. The vertically integrated, wintertime eddy heat flux is \( c_p g^{-1} \int \int v' T' \, dx \, dp \approx 6 \) PW (Peixóto and Oort 1984), where \( c_p \) is the specific heat capacity at constant pressure, \( v \) the meridional velocity, \( T \) the temperature, \( x \) longitudinal distance and \( p \) the pressure. The prime denotes a departure from the zonal mean.

(b) The meridional mass flux, \( \mathcal{V} \). When viewed in isentropic coordinates the transfer of thermal energy polewards is effected by a poleward mass flux in the upper troposphere and lower stratosphere, as analysed by Johnson (1988), Iwasaki (1989), Jackes et al. (1994) and Bartels et al. (1998). This representation has the advantage of giving a clearer (or at least alternative) picture of how the air masses are moving. The strength of this circulation is about 100 Tg s\(^{-1}\).

(c) The meridional eddy flux of potential vorticity. This may be expressed as an isentropic flux of Ertel's potential vorticity, or, an isobaric flux of quasi-geostrophic potential vorticity. Potential vorticity plays a key role in the dynamics of the extratropical flow. In this study it also provides a demarcation between tropospheric and stratospheric air. Bartels et al. (1998) calculate the value of the meridional eddy flux of potential vorticity, from ECMWF analyses, at the wintertime tropopause to be \( \sigma v^* \mathcal{P}^*_s \Big|_{x, t} \approx 3 \times 10^{-5} \) m s\(^{-2}\), where the * denotes a departure from the mass-weighted isentropic mean, \( \sigma \) is the pseudo density in isentropic coordinates, and \( \mathcal{P} \) is Ertel's potential vorticity. The superscripts \( x \) and \( t \) indicate the variables averaged over and
the subscripts \( y \) and \( \theta \) indicate the variables held constant (\( \theta \) is potential temperature). Charney and Stern (1962) show that the isentropic gradient of Ertel's potential vorticity multiplied by the pseudo density is approximately equal to the isobaric gradient of the quasi-geostrophic potential. Multiplying the unnumbered equation following Eq. (2.31) in their paper by the pseudo density, using conservation of mass to transform the equation into flux form, averaging over time and the zonal coordinate, and finally performing a partial integration with respect to the meridional coordinate, gives

\[
\overline{\sigma v^* \overline{P}^* |_{y, \theta}^{x,t}} \approx \overline{v^* |_{y, \rho}^{x,t}}
\]

where \( q \) is the quasi-geostrophic potential vorticity.

(d) The meridional eddy flux of tropopause potential-temperature anomalies. It is convenient to quantify the movement of the tropopause through the evolution of its potential temperature, \( \theta_{tp} \), since the latter is conserved in adiabatic motion. The meridional flux of this quantity then provides a measure of meridional transport processes at the tropopause level. Its amplitude was estimated in J97, from theoretical considerations, as about 60 K m s\(^{-1}\).

(e) The mass flux through the tropopause on the polewards side of the jet, \( \mathcal{W}_{tp} \). The four quantities listed under (a) to (d) are primarily associated with meridional transport. It is the purpose of this study to establish a link with the vertical transport through the tropopause. The magnitude of this mass flux through the tropopause was estimated in J97 as \( \mathcal{W}_{tp} \approx 22 \) Tg s\(^{-1}\), directed upwards.
(b) Links between different aspects of the meridional circulation

The five quantitative measures of the meridional circulation listed under (a) to (e) are linked to each other through the following theoretical concepts:

(i) The transformed Eulerian mean (TEM) theory of Andrews and McIntyre (1976) relates eddy heat fluxes to the mean meridional circulation. This theory was a major contribution to clarifying the link between eddy forcing and meridional transport.

(ii) In an extension of the TEM formulation, Tung (1986) showed how the eddy fluxes of heat and momentum were related to the eddy flux of potential vorticity.

(iii) The zonally averaged zonal-momentum equation in isentropic coordinates provides a direct link between the mean eddy flux of potential vorticity and the mean meridional mass flux (Andrews 1983; Juckes et al. 1994).

(iv) Quasi-geostrophic tropopause theory (Juckes 1994) provides a link between isobaric potential-temperature anomalies in the troposphere and potential-temperature anomalies on the tropopause.

(v) The QGT theory also provides a theoretical relationship between the potential-temperature anomaly on the tropopause and the isobaric quasi-geostrophic potential-vorticity anomaly. This allows the more traditional description of meridional motions to be related to the dynamics of the tropopause itself.

(vi) A further contribution of QGT theory is a relation between $\theta'_p$ and tropopause-height anomalies. The eddy flux of these anomalies constitutes a meridional mass flux in the lower stratosphere which must be balanced by a mass transfer through the tropopause. J97 used the links (v) and (vi) implied by QGT theory together with (iii) to show that the mass transfer through the tropopause is an intrinsic part of the meridional circulation.

(vii) The link between $\mathcal{V}$ and $\mathcal{W}_p$ mentioned in the previous paragraph relied on quasi-geostrophic scaling being valid throughout the flow. The results described in sections 3 and 4 establish a direct link which requires only that the Rossby number be small at the edge of a control volume containing the extra-tropical tropopause. The predicted mass flux is the same, but the more direct derivation requires less restrictive assumptions.

(viii) The link between $\overline{v'q'}$ and $\mathcal{W}_p$ could be regarded as a representation transformation (rtr). That is, we have a dataset of points, each with three relevant quantities: ($v'$, $q'$, $p$). The relationships between these three quantities can be represented as either a correlation between $v'$ and $q'$ at constant $p$ or as a correlation between $v'$ and $p'$ at constant $q$. The TEM theory can be viewed as a special case of a transformation between two such representations.

3. Control Volumes

An integral conservation property of potential vorticity can be exploited by considering the fluxes across the boundaries of a control volume with isentropic upper and lower surfaces enclosed by a vertical lateral boundary $S_L$ (see Fig. 2). The upper surface $S_u$ is taken to be in the stratosphere and the lower surface $S_t$ in the troposphere. The section of tropopause enclosed in the control volume will be referred to as $S_{tp}$. The results derived in sections 4 and 5 later rely on quasi-geostrophic scaling being applicable at the lateral boundaries, no restriction is imposed on the nature of the dynamics within the control volume. It will also be assumed that the tendencies both of the circulation $\int \mathbf{u} \cdot dl$ around the lateral boundary and of the enclosed mass are small.
A conceptual separation between those processes which maintain the sharpness of the tropopause and those processes which transport mass across the tropopause is implicit in the following discussion. Generally, the same physical process will contribute to both parts of this conceptual pair, but the conceptual separation can be justified by the fact that it produces a meaningful result. That is, if we assume that the tropopause remains sharp it is possible to analyse the mass flux across the tropopause without explicit reference to the processes which maintain that sharpness. The properties of these processes are of course also of some interest, but are not the subject of this note.

The analysis makes use of Ertel’s potential vorticity:

$$\mathcal{P} = \rho^{-1} (2\Omega + \zeta) \cdot \nabla \theta$$

(Ertel 1942), where $\rho$ is the density, $\zeta$ the relative vorticity and $\Omega$ the planetary rotation vector.

For present purposes a flux form of the potential-vorticity evolution equation is useful, a slightly modified form of Eq. (4.3) of Haynes and McIntyre (1987). The modification is derived by splitting the velocity into an isentropic component and a component associated with the mass flux through isentropes. Let

$$u^x = \frac{\dot{\theta} \nabla \theta}{|\nabla \theta|^2},$$

where $\dot{\theta} = D\theta/Dt$ and $D/Dt = \partial/\partial t + \mathbf{u} \cdot \nabla$ is the material derivative following the three dimensional wind vector $\mathbf{u}$. It follows directly from the definition that $u^x \cdot \nabla \theta = \dot{\theta}$ and hence

$$\left( \frac{\partial}{\partial t} + \mathbf{u}|_\theta \cdot \nabla \right) \theta = 0,$$

where $\mathbf{u}|_\theta = \mathbf{u} - u^x$. That is, $\mathbf{u}|_\theta$ is the velocity projected onto the moving isentropic surface. Note that this equation is an exact kinematic consequence of the definition of $\mathbf{u}|_\theta$ and does not involve any assumption about the value of $\dot{\theta}$. 
Substituting this decomposition of the velocity into Eq. (4.3) of Haynes and McIn-tyre (1987) leads to the following flux form of the potential-vorticity evolution equation:

$$\frac{\partial (\rho \mathcal{P})}{\partial t} + \nabla \cdot (\rho \mathbf{u}|_{\theta} \mathcal{P} - \zeta_{a}|_{\theta} \hat{\theta} - \mathbf{F} \times \nabla \theta) = 0,$$

(1)

where \( \mathbf{F} \) represents non-conservative terms in the momentum equations (\( \mathbf{F} \) has units of acceleration) and

$$\zeta_{a}|_{\theta} = \zeta_{a} - \frac{(\zeta_{a} \cdot \nabla \theta) \nabla \theta}{|\nabla \theta|^{2}}$$

is the absolute vorticity projected onto the isentropic surface. Equation (1) is valid without approximation for non-hydrostatic flow. If (1) is put into isentropic coordinates, making use of the hydrostatic approximation, Eqs. (2.4–2.7) of Haynes and McIntyre (1987) are recovered. The fact that the tendency of \( \rho \mathcal{P} \) can be written, without approximation, as the divergence of a vector follows from the fact that \( \rho \mathcal{P} \) is itself the divergence of a vector: \( \rho \mathcal{P} = \nabla \cdot [(2\mathbf{Q} + \zeta) \theta] = \nabla \cdot [(\mathbf{Q} \times \mathbf{r} + \mathbf{u}) \times \nabla \theta] \), where \( \mathbf{r} \) is the position vector. The calculation below will make use of the fact that the second and third terms of the flux in (1) are perpendicular to \( \nabla \theta \). The diabatic circulation does not appear explicitly in the first term of the flux, but it is present in the divergent component of \( \mathbf{u}|_{\theta} \). In fact, Haynes and McIntyre show that this is the dominant diabatic contribution to the hydrostatic form of this equation on the planetary scales, with the second term being an order of Rossby number smaller.

Integrating (1) over the control volume and over a time interval \( t_{0} \rightarrow t_{1} \) gives

$$\int_{t_{0}}^{t_{1}} \int_{S_{ht}} \rho \mathcal{P} \mathbf{u}|_{\theta} \cdot \mathbf{n} \, dA \, dt = \int_{t_{0}}^{t_{1}} \int_{S_{ht}} (\zeta_{a}|_{\theta} \hat{\theta} + \mathbf{F} \times \nabla \theta) \cdot \mathbf{n} \, dA \, dt$$

(2a)

$$- \left( \mathbf{n} \left[ (\mathbf{Q} \times \mathbf{r} + \mathbf{u}) \times \nabla \theta \right] \cdot \mathbf{n} \, dA \right)_{t_{0}}^{t_{1}},$$

(2b)

where \( \mathbf{n} \) is the unit outward normal to \( S_{ht} \). Because of the divergence form of (1) all terms are evaluated at the boundary. The last term makes use of the divergence expression for \( \rho \mathcal{P} \) given in the last paragraph. There is no advective flux through the isentropic upper and lower boundaries because \( \mathbf{u}|_{\theta} \) is projected onto the moving isentropic surfaces. The non-conservative terms do not have any contribution from these surfaces because they are perpendicular to \( \nabla \theta \).

Equation (2) is still an exact result, but we can now estimate the various terms and neglect the smaller ones. As in J97 the lateral boundary of the control volume will be taken as a circle of constant latitude close to the centre of the storm tracks as defined by the meridional eddy heat flux. Previous diagnostic studies give \( \nabla = \int_{S_{ht}} \rho \mathbf{u}|_{\theta} \cdot \mathbf{n} \, dA \approx 10^{11} \) kg s\(^{-1}\). Taking a typical tropospheric potential-vorticity value of \( 10^{-6} \) m\(^2\)s\(^{-1}\) K kg\(^{-1}\) and a time period, \( t_{1} - t_{0} \), of one month (about \( 2.5 \times 10^{6} \) s) gives \( 2.5 \times 10^{11} \) K m\(^2\)s\(^{-1}\) for the left-hand side of (2a). For the diabatic term we can take \( |\zeta_{a}|_{\theta} \approx \left| \frac{\partial u}{\partial z} \right| \approx 5 \times 10^{-3} \) s\(^{-1}\) and \( \hat{\theta} \approx 1 \) K day\(^{-1}\). If the lateral boundary is taken to be 8 km deep and 20 000 km in circumference the integral is \( 2.5 \times 10^{10} \) K m\(^2\)s\(^{-1}\). For (2b), which is the change in circulation integrated over the potential-temperature range spanning the control volume, we can estimate the change in \( u \) as 10 m s\(^{-1}\) and the \( |\nabla \theta| \) as \( 3 \times 10^{-3} \) K m\(^{-1}\), which gives \( 4.8 \times 10^{9} \) K m\(^2\)s\(^{-1}\) for the change in the integral. The two terms on the right-hand side are clearly much smaller than the scale estimate of the left-hand side.
The final assumption is that the variation in $\mathcal{P}$ is dominated by the sharp contrast between stratosphere and troposphere, so that its value in the two regions can be approximated by two constants, $\mathcal{P}_t$ and $\mathcal{P}_s$ respectively. It then follows that

$$\mathcal{P}_t \int_{S_{ht,t}}^{t_0,t_1} \rho \mathbf{u} \cdot \mathbf{n} \, d\mathcal{A} + \mathcal{P}_s \int_{S_{ht,s}}^{t_0,t_1} \rho \mathbf{u} \cdot \mathbf{n} \, d\mathcal{A} \approx 0, \quad (3)$$

where $S_{ht,t}$ and $S_{ht,s}$ are, respectively, the tropospheric and stratospheric portions of $S_{ht}$ and

$$\int_{S_{ht,t}}^{t_0,t_1} () \, d\mathcal{A} = \int_{t_0}^{t_1} \int_{S_{ht,t}} () \, d\mathcal{A} \, dt.$$

In (3) the approximation $\mathbf{u} \cdot \mathbf{n} \approx \mathbf{u}|_{\theta} \cdot \mathbf{n}$ has been made, since the difference between the two velocity vectors is firstly small and secondly almost perpendicular to $\mathbf{n}$. This approximation is consistent with the neglect of the right-hand side in (2).

The accuracy of the assumption of uniform potential vorticity in the stratosphere could be improved by using an alternative form of the potential vorticity with less vertical variation in the stratosphere, as proposed by Lait (1994). This would involve replacing $\theta$ with $\Theta = \frac{\theta_{00}^{1+1/k}}{\theta^{1/k}}$ throughout the analysis, where $\theta_{00}$ is a reference potential temperature and $\kappa = (c_p - c_v)/c_p$. For the sake of simplicity we remain with Ertel’s potential vorticity here.

Conservation of mass in, respectively, the tropospheric and stratospheric portions of the control volume gives

$$\int_{S_{ht,t}}^{t_0,t_1} \rho \mathbf{u} \cdot \mathbf{n} \, d\mathcal{A} + \int_{S_{tp}}^{t_0,t_1} w_{tp} \rho \, d\mathcal{A} - \int_{S_{lt}} w_{l} \rho \, d\mathcal{A} = 0 \quad (4a)$$

and

$$\int_{S_{ht,s}}^{t_0,t_1} \rho \mathbf{u} \cdot \mathbf{n} \, d\mathcal{A} - \int_{S_{tp}}^{t_0,t_1} w_{tp} \rho \, d\mathcal{A} + \int_{S_{ls}} w_{s} \rho \, d\mathcal{A} = 0, \quad (4b)$$

where $w_l$, $w_s$ and $w_{tp}$ are the relative velocities perpendicular to the lower, upper and tropopause control surfaces respectively (see Fig. 2). That is, $w_{tp} = \mathcal{O}/|\nabla \mathcal{P}|$ and $w_{lt,s} = \mathcal{O}/|\nabla \theta|$. The term corresponding to the change in mass of the control volume has been neglected. This neglect can be justified by considering the mass flux through the lateral boundary integrated over one month, which comes to about $2.5 \times 10^{17}$ kg. This is equivalent to a layer of air 300 mb deep spread over the area north of 45°N. The mean pressure tendencies are considerably smaller than 300 mb/month and so may be neglected here.

Equations (3) and (4) can now be used to derive diagnostic relationships between the mass fluxes below, above, and through, the tropopause.

4. The Meridional Mass Flux

A key aspect of J97 is the prediction of equatorward motion in the lowermost stratosphere. This result can be derived by applying the results of section 3 to a control volume bounded by a circle of constant latitude. Suppose the net meridional mass flux integrated over these isentropes, $\mathcal{V}$, is given by

$$\mathcal{V} = \mathcal{V}_t + \mathcal{V}_s, \quad (5)$$
where the subscripts indicate the tropospheric and stratospheric parts of the flux. In terms of the integrals used above

\[ V_t = \int_{S_{t,t}} \rho \mathbf{u} \cdot \mathbf{n} \, dA \quad \text{and} \quad V_s = \int_{S_{t,s}} \rho \mathbf{u} \cdot \mathbf{n} \, dA. \]

The potential-vorticity budget (3) requires

\[ V_t P_t + V_s P_s = 0. \quad (6) \]

Combining (5) and (6) gives

\[ V_t = \sqrt{P_s - P_t} \quad \text{and} \quad V_s = -\sqrt{P_s - P_t}. \]

The mass flux in the stratosphere is thus negative, i.e. directed towards the equator. This is equivalent to a result of J97. There, however, the derivation required that the Rossby number be small over the entire flow. The result here only requires that it be small at the lateral control surface bounding the region of interest. The value of \( V \) is about 100 Tg s\(^{-1}\) (Johnson 1988; Iwasaki 1989; Luckes et al. 1994; Bartels et al. 1998). Taking \( P_s = 4P_t \) then implies an equatorward mass flux in the lowermost stratosphere of magnitude 33 Tg s\(^{-1}\). This flux is of the order of magnitude suggested by Brewer, but in the opposite direction. We will return to this point in section 6: the next section considers the mass flux through the tropopause which is implied by the results of this section.

5. Mass Flux Through the Tropopause

The meridional mass flux derived in section 4 implies a net transfer of air from the troposphere into the stratosphere at high latitudes and back into the troposphere at lower latitudes. As indicated in the introduction this zonal mean flux is not the same as a typical parcel motion. One alternative to the zonal mean, discussed in more detail by J97, is an average taken over parcels on the tropopause with a given potential temperature. The theory indicates a net transfer to the troposphere for the average over parcels with the low tropopause potential temperatures typical of high latitudes. The zonal mean result, however, implies that this transfer must take place at low latitudes. The two conclusions are consistent if the transfer is associated with parcels of high latitude air which have been displaced to low latitudes.

Here another problem will be considered: the observed dryness of the lowermost stratosphere would appear to be incompatible with large-scale rising motion through the extra-tropical tropopause. Before addressing this point a result from J97, relating heating rates to tropopause exchange, is re-derived by considering the budget of Ertel’s potential vorticity within a control volume spanning the tropopause.

We now consider a control volume over a relatively small area, so that the upper and lower isentropic boundaries (\( S_u \) and \( S_l \) respectively) are close to the tropopause (\( S_{tp} \) and the rates of mass transfer across these boundaries are almost uniform. As before, \( w_u \) and \( w_l \) are the velocities across the upper and lower surfaces (i.e. the rate of mass transfer divided by area and density), and \( w_{tp} \) is the velocity across the tropopause within the control volume (see Fig. 2). In this configuration the justification for the neglect of the circulation tendency is less clear. As before there is cancellation between diabatically induced tendencies and adiabatically induced tendencies, the latter tending to create anomalies by quasi-horizontal advection and the former tending to dissipate
anomalies. In the long term these two components of the system balance out so that the anomaly variance is maintained. The analysis below considers those diabatic effects that are necessary to cancel adiabatic processes that generate the tropopause potential-temperature anomaly.

From (4) it follows that the mass fluxes \( \mathcal{V}_s \) and \( \mathcal{V}_t \) across the lateral boundaries satisfy:

\[
\mathcal{V}_s + (w_{\text{tp}} - w_s) A \rho = 0, \quad (7a)
\]

\[
\mathcal{V}_t + (w_t - w_{\text{tp}}) A \rho = 0. \quad (7b)
\]

Combining (6) and (7) gives

\[
(w_s - w_{\text{tp}}) \mathcal{P}_s + (w_{\text{tp}} - w_t) \mathcal{P}_t = 0.
\]

Solving for the velocity through the tropopause gives:

\[
w_{\text{tp}} = \frac{w_s \mathcal{P}_s - w_t \mathcal{P}_t}{\mathcal{P}_s - \mathcal{P}_t}. \quad (8)
\]

In the low Rossby-number limit \( \mathcal{P} \approx f \Gamma \), giving

\[
w_{\text{tp}} = \frac{w_s \Gamma - w_t \Gamma_t}{\Gamma - \Gamma_t}.
\]

In order to arrive at the result found in J97 a further approximation is needed, namely \( w_{\text{tp}} \Gamma_s / \Gamma_t \approx \dot{\theta}/f \) where \( \Gamma = \partial \theta / \partial z \) (this amounts to approximating \( |\nabla \theta| \) with \( \partial \theta / \partial z \)). Note that the velocities \( w_{\text{tp}} \) are vertical velocities relative to the corresponding isentropic control surfaces. It then follows that

\[
w_{\text{tp}} = \frac{\dot{\theta}_s - \dot{\theta}_t}{\Gamma_s - \Gamma_t} \quad (9)
\]

as derived in J97. That is, this result relies on the Rossby number being small on the boundaries of the control volume. It does not depend on the processes that effect the mass flux across the tropopause. It is possible to derive a diagnostic result in this form because we are considering a quasi-steady state, that is a state in which there may be significant variability but for which the tendency terms integrated over the volume are small. In this quasi-steady state the problem is over-determined and can be solved without referring to all the participating processes.

The rate of change of tropopause potential temperature is given by

\[
\dot{\theta}_{\text{tp}} = \dot{\theta} - w_{\text{tp}} \Gamma,
\]

where \( \dot{\theta} \) is the rate of change of \( \theta \) following a fluid trajectory and \( \dot{\theta}_{\text{tp}} \) is the corresponding rate of change for a trajectory projected onto the tropopause. On the right-hand side \( \dot{\theta} \) and \( \Gamma \) are written without subscripts because this relation must hold both just above and just below the tropopause. It is readily verified that using either the tropospheric or stratospheric values of \( \dot{\theta} \) and \( \Gamma \) and (9) for \( w_{\text{tp}} \) one obtains the same result:

\[
\dot{\theta}_{\text{tp}} = \frac{\Gamma_s \dot{\theta}_s - \Gamma_t \dot{\theta}_t}{\Gamma_s - \Gamma_t}.
\]

Typically, on synoptic scales, a warm tropospheric anomaly is associated with a cold stratospheric anomaly (e.g. Hoskins 1991). If it is assumed that diabatic effects generally
act to damp anomalies, this would imply a tendency for the two terms in the numerator, \( \Gamma_s \dot{\theta}_t \) and \( -\Gamma_t \dot{\theta}_s \), to reinforce each other. There will not, in reality, be a sudden change of sign in the diabatic cooling rates at the tropopause, but we might generally expect to find a gradient there such that \( |\dot{\theta}_t| > |\dot{\theta}_s| \). In either case, it follows that

\[
|\dot{\theta}_{tp}| = \left| \dot{\theta}_t + \frac{\Gamma_t (\dot{\theta}_t - \dot{\theta}_s)}{\Gamma_s - \Gamma_t} \right| > |\dot{\theta}_t|.
\]

This shows that although the air in a high-latitude warm anomaly may be sinking due to radiative effects it is likely to be overtaken by the tropopause which is, due to the same adiabatic processes, sinking faster.

It is, of course, a severe simplification to represent heating rates by single stratospheric and single tropospheric values. Zierl and Wirth (1997) have carried out a more detailed calculation with an axi-symmetric vortex and a realistic representation of radiation. In their model of a decaying anticyclone the air immediately below the perturbed tropopause was not sinking but rather was more or less stationary. As here, the transfer of air from troposphere to stratosphere took place primarily through the downwards motion of the tropopause.

Figure 3 summarizes the arguments and results of this section schematically. On the left an anomalously high tropopause is depicted, with a cold stratospheric anomaly overlying a warm tropospheric anomaly. If the anomalies are dissipated by non-conservative processes there will then be diabatic cooling in the troposphere and warming in the stratosphere, indicated by the diverging arrows. This divergence leads to a positive potential-vorticity tendency which implies a transfer of mass into the stratosphere. The calculation above shows that the tropopause actually descends faster than the air in the upper troposphere, so as it sinks the latter is transferred into the stratosphere.
6. The Brewer-Dobson Circulation

The results derived in J97 and above raise some interesting questions about the Brewer-Dobson circulation. Dobson et al. (1946) measured extremely dry air in the mid-latitude lower stratosphere and concluded that the air must have been freeze dried by passing through a region of extremely low temperature. In the zonal-mean climatology the only part of the atmosphere with sufficiently low temperatures is the tropical tropopause. This led Brewer (1949) to propose a circulation rising in the tropics and flowing out to mid latitudes in the lower stratosphere. He suggested that this outflow should occupy a layer extending several kilometres above the tropopause and estimated a mass flux of about 30 Tg s\(^{-1}\), based on an estimate of the radiative cooling rate as 0.5 K day\(^{-1}\).

Subsequent research has clarified the nature of the circulation rising in the tropics. The absolute angular-momentum budget plays a key role. The air in the tropics has a very high absolute angular momentum and this must be discarded as the air moves polewards. The language here has become somewhat controversial: some would see the poleward movement of the air as a consequence of processes removing its angular momentum, others would see the removal of its angular momentum as a consequence of its poleward motion. We can bypass this discussion if we concentrate on the diagnostic issues and resist the temptation to explain everything in one go. Andrews and McIntyre (1976) have shown that the meridional motion of the mid and upper troposphere can be described elegantly with the help of the residual mean circulation. They built on an earlier result (Eliassen and Palm 1961) showing how the eddy forcing of the mean motion, which is comprised of both meridional eddy momentum fluxes and meridional eddy heat fluxes, could be expressed in a single quantity, now known as the divergence of the Eliassen-Palm flux. The stream function of the residual mean circulation (e.g. Andrews et al. 1987) is given by

\[ \psi^* = \psi + \psi_{v\theta}, \quad \text{where} \quad \psi_{v\theta} = \frac{\rho v \theta ^{x,t}}{\partial \theta ^{x,t}/\partial z} \]

and \( \psi \) is the zonal-mean mass stream function, defined such that \( \partial \psi / \partial z = -\rho \bar{v}^{x,t} \). The bar denotes an average over the coordinates indicated in the superscript. The primed variables are departures from the zonal mean. In forming the time mean it has been assumed that the temporal fluctuations in \( \partial \theta ^{x,t}/\partial z \) are small: this condition is generally satisfied in the middle and upper stratosphere but not necessarily in the lowermost stratosphere. The residual mean circulation gives an estimate of the meridional mass circulation in isentropic coordinates (Dunkerton 1978; Appenzeller et al. 1996).

Figure 4 shows this circulation calculated from eddy momentum fluxes supplied by Randel (1987). These fluxes have in turn been derived from satellite radiance measurements in the stratosphere and meteorological analyses below 100 hPa. The quantities plotted are the mass flux, \( \Psi^* = 2\pi r\psi^* \), where \( r \) is the Earth's radius, and \( \Psi_{v\theta} = 2\pi r\psi_{v\theta} \). The mean meridional velocity is estimated from the mean zonal-momentum equation which, to leading order in Rossby number, is

\[ f \bar{v}^{x,t} = \frac{\partial \bar{u} \bar{v}^{x,t}}{\partial y}. \]

The first attempt to evaluate the mean circulation in the stratosphere in this way was made by Gilman (1965), who used a more complex expression accounting for advection of relative momentum by the mean velocity. The equation must then be
solved by integrating along characteristics which are the lines of constant absolute angular momentum. The errors incurred by using the simplified expression above are considerably smaller than the unavoidable errors which occur when the eddy momentum flux is calculated from balanced winds (Boville 1987). The contour interval in Fig. 4 is irregular (see caption) in order to show the structure of the circulation in the upper atmosphere where the magnitude of the mass flux is relatively weak. Equation (11) is used for latitudes outside the band from 10°S to 10°N. Inside this band it is assumed that $\psi$ changes linearly (this is equivalent to prescribing a uniform vertical velocity across this tropical band). Horizontal and vertical components of the arrows show the meridional velocity in the transformed Eulerian mean and the cooling rate estimated from the transformed Eulerian-mean theory respectively.

Figure 4(b) shows the contribution the heat flux alone makes to the mean meridional circulation, $\Psi_{\theta\theta}$. The amplitude is similar to the total, $\Psi^*$, shown in Fig. 4(a), but the pattern is displaced polewards. The uncertainty in $\Psi^*$ comes mostly from the meridional momentum flux used to calculate $\Psi$ (Boville 1987), the heat-flux contribution (Fig. 4(b)) is relatively well determined. Adding $\Psi$ to $\Psi_{\theta\theta}$ causes a shift of the circulation without much change in magnitude. This means that the mass flux across a particular pressure level can be estimated directly from the heat fluxes at that level. For instance, if $\partial\theta/\partial z$ at 100 hPa is 12 K km$^{-1}$, the mass flux corresponding to a temperature flux of 20 K m s$^{-1}$ would be $\Psi_{\theta\theta} = 6.6$ Tg s$^{-1}$, which is close to the total depicted in Fig. 4. This shows that the value is robust: there are questions about the accuracy of fluxes in the middle and upper stratosphere, but the value of $\Psi_{\theta\theta}$ at 100 hPa is reliable.

The calculation of $\Psi^*$ for Fig. 4 did not account for the eddy fluxes above 1 hPa. In the mesosphere, breaking gravity waves become the dominant agent of angular-momentum transport, but they are only associated with a mass flux of around 0.2 Tg s$^{-1}$. This value is obtained from Garcia and Boville (1994), who estimate a descent velocity, due to gravity-wave forcing, of about 0.4 mm s$^{-1}$ at 5 hPa: if this covers half the hemisphere the downwards mass flux would be the figure given above – the precise figure depends on the height chosen as reference level and also, of course, on the estimated amplitude of the gravity waves, but the figure given should be of the right order of magnitude. There is some confirmation in that the adiabatic heating associated with this sinking improves the agreement of Garcia and Boville’s model with observations.

Rapid meridional transport occurs in the upper winter stratosphere, where large amplitude planetary waves generate intense eddy fluxes of momentum. Although the meridional transport here is fast it contributes only a fraction of the total mass transport, about 2 Tg s$^{-1}$, because of the low density. The height of the boundary between the upper stratospheric region, where non-linear dissipation dominates, and the lower region, where radiative dissipation is larger, has not been systematically analysed, but an estimate of about 20–50 hPa appears reasonable. The time-scales associated with radiative dissipation in the lower stratosphere are significantly longer than those associated with the nonlinear dynamics of the upper stratosphere, but the mass-weighted amplitude of the waves is greater, so that the net mass flux associated with this process is larger. This region of the stratosphere accounts for a mass flux of about 3 Tg s$^{-1}$. The total mass flux rising through the tropical tropopause, according to this dataset, is about 5 Tg s$^{-1}$. Rosenlof and Holton (1993) quote a slightly larger value of almost 8 Tg s$^{-1}$. This mass flux is substantially less than the amount of descent inferred by Brewer. The radiative budget of the lowermost stratosphere is difficult to analyse, but it appears unlikely that the wintertime cooling rates are substantially less than the 0.5 K day$^{-1}$.
Figure 4. The mass stream function of the transformed Eulerian-mean meridional circulation: (a) $\Psi^*$ and (b) $\Psi_{\theta\theta}$. Thick contours are plotted at $\pm 0.1$, $\pm 1$, $\pm 10$ and 100 Tg s$^{-1}$, thinner contours are at two and five times these powers of ten. The arrows represent the motion in the meridional plane. The horizontal component of the vector is proportional to the residual mean meridional velocity and the vertical component proportional to the mean heating rate estimated as the residual mean vertical motion times $\partial \bar{\theta} / \partial z$. 
which led to Brewer's mass-flux estimate. In other words, Brewer's estimate requires another mass source and this mass source could be provided by the transfers associated with baroclinic waves.

Although Brewer's mass-flux estimate was based on cooling rates, it was motivated by the extreme dryness which Dobson et al. (1946) had measured in the lowermost stratosphere. This dryness, he suggested, implied that the air had passed through the tropical cold trap. An alternative, coming from this study, is that the air enters the lowermost stratosphere through the extremely cold regions associated with the tropopause in mid-latitude high-pressure regions. The dryness of the air rules out the possibility that it rises into the stratosphere through the extra-tropical tropopause. The theory presented here suggests a subtle alternative: the air is entering the stratosphere in the extratropics, but sinking as it does so.

The mass circulation in the stratosphere has been calculated from meridional heat fluxes whereas the value used for the tropospheric circulation was quoted from analyses of the isentropic mean circulation. The value of \( \Psi_{v\theta} \) can be used to show that the two are consistent. The total meridional heat flux in the winter mid-latitude troposphere reaches about \( 6 \times 10^{15} \) W (e.g. Peixóto and Oort 1984). Approximately 80% of this is associated with sensible-heat fluxes, the rest is dominated by latent-heat fluxes. This is equivalent to a poleward temperature flux of \( 24 \) K m s\(^{-1} \) extending up to 200 mb at 45\(^o\)N. In practice the temperature fluxes are stronger near the surface than in the mid troposphere, but this value gives the order of magnitude. The corresponding mean meridional mass flux above 500 mb is \( \Psi_{v\theta} \approx 70 \) Tg s\(^{-1} \). The value obtained increases monotonically as the pressure level is moved down to 1000 mb, but its interpretation as a mass flux is only valid so long as the isentrope corresponding to the zonal-mean potential temperature does not intersect the surface at the latitude in question. The above value is consistent with diagnostic studies of the tropospheric meridional mass flux in isentropic coordinates cited earlier.

7. ISENTROPIC MIXING

Holton et al. (1995) suggested that isentropic mixing is an important mechanism for the transfer of tropospheric air into the lowermost stratosphere. It is argued, as here, that air from the subtropics is advected polewards along isentropes. It is supposed that at some stage during this northward advection it is transformed into stratospheric air. The filamentary structures which are consistent with this hypothesis are indeed observed. However, the meridional motion of air on a rapidly rotating planet is strongly restricted by the conservation of angular momentum. When air moves significant distances across the meridians it must do so in a structure which enables it to transfer its angular momentum to air moving in the opposite direction. For this reason the dominant meridional mass fluxes in the troposphere are associated with synoptic- and planetary-scale waves.

By concentrating on small-scale mixing and ignoring the implications of the angular-momentum budget Holton et al. (1995) miss a key aspect of the circulation: namely, the diagnostic link between the meridional heat flux in the troposphere and the flux of tropospheric air into the stratosphere. The meridional motion is unquestionably dominated by the isentropic component, but here, as in Green (1970), it is regarded as being primarily a transfer process distinct from mixing.

There is no direct equivalence between published results and the quantities described in this paper, but it is possible to check the consistency with the diagnostics presented in Bartels et al. (1998) by using the results of J97. The mass fluxes discussed here
imply a meridional flux of tropopause potential temperature of \( \overline{v' \theta_p} = 60 \text{ K m s}^{-1} \). The meridional circulation derived for July in the southern hemisphere by Bartels et al. (1998) is a factor of 0.6 weaker than that used to derive the above value, so the relevant value for their study would be 36 K m s\(^{-1}\). Using Eq. (39) of J97 to relate the tropopause potential-temperature anomaly to an equivalent potential-vorticity anomaly gives a vertically integrated meridional potential-vorticity flux of 0.9 m\(^2\)s\(^{-2}\). The local values of the potential-vorticity flux diagnosed by Bartels et al. (1998), multiplied by the isentropic pseudo-density, reach about \( 10^{-4} \) m s\(^{-2}\). This would be consistent if the flux extended through a depth of 9 km. A more detailed analysis would be desirable, but this preliminary estimate shows that the theory is consistent with published analyses of isentropic mixing, once allowance is made for the fact that the meridional circulation deduced by Bartels et al. (1998) is weaker than the values used in J97.

8. Discussion

In J97 a series of results produced a link between the mean meridional mass flux in the troposphere and a mass exchange between the troposphere and stratosphere. The analysis presented above has reproduced this result with a more direct derivation. Essentially, the equatorward eddy flux of potential vorticity in the upper troposphere implies that high-potential-vorticity air (i.e. stratospheric air) is, in the mean, moving equatorwards. This fact is quantified by considering the potential-vorticity and mass budgets of a control volume spanning the tropopause polewards of the latitude of the storm tracks. It is found that the implied equatorward mass flux in the stratosphere is of considerable amplitude, significantly greater than the mass flux entering the lowermost stratosphere from above.

A clearer picture of the structure of this circulation can be gained by consideration of the dynamics of the tropopause. In J97 this leads to the conclusion that polewards motion is correlated with a high tropopause and equatorward motion with a low tropopause. This means that the equatorward mass flux in the lower stratosphere actually occurs below the mean tropopause height and the poleward mass flux in the troposphere is partly above the mean tropopause height. This aspect of the circulation is sketched in Fig. 5. The sloping solid line indicates the mean position of the tropopause, and dashed lines the upper and lower limits of fluctuations in tropopause height. The upper arrow (1) indicates the poleward mass flux in the upper troposphere, which is associated with an anomalously high tropopause. At high latitudes this air cools through diabatic effects, and as it does so the tropopause descends and some of the air is transferred into the stratosphere. The lower arrow (3) indicates equatorward transport in the lower stratosphere associated with an anomalously low tropopause. At (4) this air warms. As it does so the tropopause rises and some of the air is returned into the troposphere. Tropospheric air rises as it moves polewards, but this branch of the motion is adiabatic, so does not directly induce exchange. The exchange processes are associated with cooling and hence descent.

Brewer (1949) suggested the existence of a circulation rising through the tropical tropopause and flowing polewards in a layer extending a few kilometres above the tropopause. He estimated the strength of this circulation at 30 Tg s\(^{-1}\). Later research has shown that there is indeed a circulation rising through the tropical tropopause, but it is substantially weaker than that proposed by Brewer and extends into the middle and upper stratosphere. The results presented here reinforce the conclusion of J97 that there is also a circulation in the few kilometres above the tropopause of the strength
suggested by Brewer, but this circulation enters the stratosphere at high latitudes and
flows equatorwards.

Rising air is typically at or near saturation, so one might expect a large flux of
mass from the troposphere into the stratosphere to be associated with high relative
humidities in both air masses, which would imply a clear contradiction between the
above conclusion and the observed dryness of the lowermost stratosphere. The analysis
presented earlier has shown that this need not be the case, since the mechanism for
the transfer of air from the troposphere to the stratosphere does not necessarily involve
rising motion but rather downward motion of the tropopause relative to air parcels. The
idealized calculations carried out here do not give any direct indication of the residence
times associated with tropospheric air entering the stratosphere nor of the distribution of
the suggested mass flux over the various isentropic levels. In J97, however, an estimate
of a vertical transport coefficient was made, describing the spread of tropospheric air
into the stratosphere, at 4 to 8 m²s⁻¹. This would be a maximum at the tropopause and
decay towards a much smaller value in the middle stratosphere. It is important to note
that this is vertical transport relative to the tropopause and is much larger than vertical
transport relative to isentropes. It is argued in J97 (last two paragraphs of discussion) that
there is no fundamental inconsistency between a transport coefficient of this magnitude
and the observed transition of tracer properties through the lowermost stratosphere.

This study has concentrated on the mass flux and it should, of course, be born in
mind that this is not the only measure of importance of the circulation: the smaller
mass fluxes in the upper stratosphere, for instance, are important because of their role
in transporting ozone out of the tropics. Nevertheless, the characterization of systematic
large-scale movement and transformation of air-masses is an important aim in itself.

ACKNOWLEDGEMENTS

I would like to thank G. Zängl and the anonymous reviewers for helpful criticism of
earlier manuscripts.
REFERENCES

Gilman, P. A. 1965 The mean meridional circulation of the southern hemisphere inferred from momentum and mass balance. Tellus, 17, 277–284
Haynes, P. H. and McIntyre, M. E. 1987 On the evolution of vorticity and potential vorticity in the presence of diabatic heating and frictional or other forces. J. Atmos. Sci., 44, 828–841
1999 The structure of idealised upper tropospheric shear lines. J. Atmos. Sci., 56, 2830–2845


