The angular momentum constraint on climate sensitivity and downward influence in the middle atmosphere

Article

Published Version

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

The Angular Momentum Constraint on Climate Sensitivity and Downward Influence in the Middle Atmosphere

Theodore G. Shepherd and Tiffany A. Shaw

Department of Physics, University of Toronto, Toronto, Ontario, Canada

ABSTRACT

It is shown that under reasonable assumptions, conservation of angular momentum provides a strong constraint on gravity wave drag feedbacks to radiative perturbations in the middle atmosphere. In the time mean, radiatively induced temperature perturbations above a given altitude z cannot induce changes in zonal mean wind and temperature below z through feedbacks in gravity wave drag alone (assuming an unchanged gravity wave source spectrum). Thus, despite the many uncertainties in the parameterization of gravity wave drag, the role of gravity wave drag in middle-atmosphere climate perturbations may be much more limited than its role in climate itself. This constraint limits the possibilities for downward influence from the mesosphere. In order for a gravity wave drag parameterization to respect the momentum constraint and avoid spurious downward influence, any nonzero parameterized momentum flux at a model lid must be deposited within the model domain, and there must be no zonal mean sponge layer. Examples are provided of how violation of these conditions leads to spurious downward influence. For planetary waves, the momentum constraint does not prohibit downward influence, but it limits the mechanisms by which it can occur: in the time mean, downward influence from a radiative perturbation can only arise through changes in reflection and meridional propagation properties of planetary waves.

1. Introduction

It has for some time been widely accepted that in order to provide a realistic simulation of the climate of the middle atmosphere, it is necessary to represent the angular momentum transfer and resulting torque exerted within the middle atmosphere by internal gravity waves propagating up from the troposphere (see, e.g., McLandress 1998). Models without a representation of this process suffer from the so-called cold pole problem, namely a severe negative temperature bias in the wintertime stratospheric polar vortex (Boville 1995). Such a bias is intolerable for accurate simulation of polar ozone depletion, for example (Austin et al. 2003). Because the disturbances responsible for most of the gravity wave drag are believed to have horizontal length scales well below the spatial resolution of most climate models, their effects must be parameterized.

Unfortunately, the observational constraints on gravity wave drag are insufficient to provide much of a constraint on gravity wave drag parameterizations (Fritts and Alexander 2003). This means that in practice, the parameterizations are tuned in a rather arbitrary manner to obtain reasonable mean climates. Tuning may

occur through adjustable parameters, or (for flux-based parameterizations) through specification of the gravity wave source spectrum. Since different parameterizations give different results, even with similar source spectra (C. McLandress 2003, personal communication), it seems likely that through tuning, deficiencies in one aspect of the parameterization would just be compensated for in another aspect.

A key question, then, is whether the manner in which a gravity wave drag parameterization is tuned affects the response of a climate model to a radiative perturbation, such as changes in greenhouse gas distributions, solar variability, a volcanic eruption, or ozone depletion. The physical realism of the response of a parameterization to radiatively induced changes is, of course, a general issue in climate science, for example, cloud–radiation feedback. It was previously argued by Shepherd et al. (1996) that Rayleigh drag sponge layers, which are often used in climate models as a crude representation of gravity wave drag, can lead to spurious dynamical feedbacks even when the radiative perturbations are applied below the sponge layer. In particular, the fact that Rayleigh drag violates angular momentum conservation leads to a physically spurious meridional circulation extending to the surface of the earth. Since key questions for climate change are the influence of the mesosphere on the stratosphere, and of the stratosphere on the troposphere, it is important to distinguish true from spurious downward influence.

Corresponding author address: Dr. T. G. Shepherd, Department of Physics, University of Toronto, 60 St. George St., Toronto, ON M5S 1A7, Canada.
E-mail: tgs@atmosp.physics.utoronto.ca

In this paper we show that under reasonable assumptions, conservation of angular momentum provides a potentially useful constraint on the time mean response of gravity wave drag to radiative perturbations, which strongly limits the potential for downward influence. The assumptions on the gravity waves are 1) no sensitivity of back reflection of the gravity wave spectrum to conditions in the middle atmosphere, 2) no meridional propagation of gravity waves within the middle atmosphere, and 3) fixed source strength unless the surface conditions change. These assumptions are made by most current gravity wave drag parameterizations. The assumptions on the zonal and time mean atmospheric response are 1) linearity of the wave-driven circulation (i.e., the angular momentum distribution is approximated by planetary angular momentum), 2) thermal radiative locality, and 3) that the only relaxational momentum source or sink is in the planetary boundary layer. These assumptions are those made in the theory of "downward control" (Haynes et al. 1991).

When a gravity wave drag parameterization respects the momentum constraint, there is exactly zero direct downward influence from gravity wave drag alone. By the phrase "downward influence" we mean changes below a given level z resulting from radiatively induced temperature changes above z (see section 2). According to this definition, a change in mesospheric gravity wave drag and the associated stratospheric circulation that results from tropospheric climate change is not considered to be an example of downward influence (even though the principle of downward control applies). Spurious downward influence in a climate model can arise when the momentum constraint is violated. Implications for true downward influence are discussed. These arise mainly from changes in the reflection or meridional propagation properties of planetary waves, either as a direct response to the radiative perturbation or as an indirect response to the gravity wave drag changes induced by that perturbation.

2. The wave-driven circulation and "downward control"

In the middle atmosphere, the thermodynamic (energy) balance for steady or time-averaged conditions is between adiabatic heating/cooling due to vertical motion and diabatic cooling/heating due mainly to radiation. This is represented in the quasigeostrophic (QG) approximation to the transformed Eulerian mean (TEM) formulation of the zonal mean equations (Andrews et al. 1987) as

$$S\overline{w}^* = Q \approx -r(\overline{T} - T_{rad}),\qquad(1)$$

where S is the static stability, $\overline{w}^*$ is the vertical residual (TEM) velocity, T is temperature, and the overbar denotes a zonal mean. We assume in this analysis that the diabatic cooling/heating Q can be regarded as a relaxation toward a radiative equilibrium temperature T_{rad} on

a time scale $1/r$. A similar balance holds in isentropic coordinates (Andrews et al. 1987). One can therefore think of the zonal mean temperature as being the sum of a radiatively determined component T_{rad} and a dynamically determined component

$$T_{dyn} = -\frac{S\overline{w}^*}{r},\qquad(2)$$

with $\overline{T} = T_{rad} + T_{dyn}$. In the global mean (represented by $\langle \cdot \rangle$), conservation of mass implies that $\langle \overline{w}^* \rangle = 0$ on every pressure surface. Thus, if S and r are independent of latitude, it follows that

$$\langle T \rangle = \langle T_{rad} \rangle,\qquad(3)$$

and the global mean temperature is in radiative equilibrium on every pressure surface. While S and r do vary with latitude, it seems that (3) is largely true in the middle atmosphere (Fomichev et al. 2002, Fig. 5). This constraint on the global mean is a powerful tool for attribution of observed temperature changes to changes in radiatively active constituents (Shine et al. 2003). It does not hold in the troposphere or above the middle mesosphere, where vertical heat fluxes (large scale in the troposphere, small scale or molecular in the upper mesosphere and thermosphere) play a fundamental role in the global mean energy balance.

The contribution of dynamical heating can, on the other hand, be very significant for the latitudinal variation of temperature. There is a component of dynamical heating associated with the time-dependent response of the atmosphere to seasonally varying radiative forcing, which induces a meridional circulation (Garcia 1987; Snieder and Fels 1988). However, this effect is not very significant in the monthly mean; for the most part, the transient TEM circulation is fairly close to that predicted by a steady-state assumption (Shine 1989; Beagley et al. 1997). In that case, the thermodynamic balance is well approximated by (1), while the angular momentum balance is between the Coriolis torque due to mean meridional motion and the angular momentum deposition (wave drag) due to the dissipation of waves propagating up from the troposphere. In the QG approximation to the TEM formulation of the zonal mean equations this latter balance takes the form

$$-2\Omega \sin\phi \overline{v}^* = \mathcal{F},\qquad(4)$$

where Ω is the rotation rate of the earth, ϕ is latitude, $\overline{v}^*$ is the meridional residual velocity, and $\mathcal{F}$ is the force per unit mass associated with the angular momentum deposition. [The angular momentum balance is actually $\rho a \cos\phi$ times (4), where ρ is density and a is the radius of the earth.] As with (1), a similar balance holds in isentropic coordinates. Using the mass continuity equation

$$\frac{1}{a\cos\phi}\frac{\partial}{\partial\phi}(\cos\phi\overline{v}^*) + \frac{1}{\rho}\frac{\partial}{\partial z}(\rho\overline{w}^*) = 0,\qquad(5)$$

where z is log-pressure altitude, it follows (under the assumption $\rho\overline{w}^* \to 0$ as $z \to \infty$) that at any altitude z,

$$\overline{w}^* = -\frac{1}{a\rho\,\cos\phi}\frac{\partial}{\partial\phi}\left(\int_z^\infty \frac{\rho\mathcal{F}\cos\phi}{2\Omega\,\sin\phi}\,dz\right). \qquad (6)$$

That is, the upwelling or downwelling $\overline{w}^*$ at any altitude z is determined by the wave drag *above* z. The relation (6) is referred to as ''downward control'' (Haynes et al. 1991).

The above considerations require careful interpretation in the Tropics. In order to satisfy a steady-state balance, (6) requires that $\mathcal{F}$ must go to zero like ϕ as $\phi \to 0$. This restriction is not a result of the QG approximation; more generally, in a steady state $\mathcal{F}$ must go to zero at the angular momentum maximum. In a time average, nonlinear effects can be important in the Tropics (Semeniuk and Shepherd 2001) and there is no simple diagnostic constraint on the local $\overline{w}^*$. A useful constraint on tropical $\overline{w}^*$ can nevertheless be obtained by integrating (6) (weighted by $\cos\phi$) in latitude; the mean tropical upwelling is determined by the wave drag at the edge of the Tropics (essentially, by the meridional mass flux into or out of the Tropics). This constraint is the ''extratropical pump'' (Holton et al. 1995).

The recognition that wave drag determines $\overline{w}^*$ and hence T_{dyn} explains much of the observed structure of the extratropical middle atmosphere (see, e.g., Shepherd 2000). It explains, for example, why the wintertime stratospheric polar vortices are warmer and weaker than expected from radiative equilibrium, and more so in the Northern than in the Southern Hemisphere. The explanation comes from the fact that planetary Rossby waves exert a negative (westward) torque where they dissipate, which from (4) drives poleward flow and downwelling over the poles. Furthermore, stratospheric planetary wave drag is restricted to the winter half of the year, when eastward zonal flow permits stationary planetary wave propagation into the stratosphere, and is stronger in the Northern than in the Southern Hemisphere. In the upper mesosphere, the observed reversal of the meridional temperature gradient at solstice, with highest temperatures over the winter pole and lowest temperatures over the summer pole, can only be the result of a wave-driven circulation that is believed to be due to momentum deposition from breaking gravity waves acting to decelerate the zonal flow in the mesosphere and lower thermosphere.

Although the meridional circulation in the middle atmosphere is, thus, mechanically driven and thermally damped, this is not necessarily evident in conventional energy diagnostics. However, this causality can be properly represented by introducing an appropriate generalization of available potential energy (Codoban and Shepherd 2003).

One may get the impression from (6) that, for example, the mesosphere affects the stratosphere through the meridional circulation induced by mesospheric wave drag. But downward control is not necessarily downward influence. Certainly, wave drag exerted within the mesosphere affects the stratosphere through the residual circulation (Garcia and Boville 1994). However, these waves have come from below; if waves transfer angular momentum into the middle atmosphere, then this angular momentum transfer necessarily drives a component of the meridional circulation, irrespective of where the waves break. For this reason, in spite of downward control, changes in mesospheric wave drag should not necessarily be regarded as a mesospheric influence on climate. An unambiguous example of downward influence would be a radiative perturbation in temperature restricted above a given level z (e.g., from solar variability or ozone loss) that led to circulation and temperature changes below z through changes in $\overline{w}^*$. Such changes would have to arise from changes in wave drag above z resulting directly from wind changes above z, and need to be distinguished from changes in tropospheric wave sources. It is important to identify real downward influence, and to ensure that climate models do not introduce spurious downward influence through their formulation of physical parameterizations. In this respect, conservation of angular momentum imposes a key constraint.

3. A constraint on radiatively induced feedbacks in gravity wave drag

In the real atmosphere, $\mathcal{F} = \boldsymbol{\nabla} \cdot \mathbf{F}/(a\rho\,\cos\phi)$ where $\mathbf{F}$ is the Eliassen–Palm (EP) flux (Andrews et al. 1987). [Divergence of EP flux is actually convergence of angular (pseudo) momentum flux (Shepherd 2003).] We assume in this analysis that for gravity waves in the middle atmosphere, the EP flux is strictly vertical and its vertical component $F^{(z)} = -a\rho(\cos\phi)\overline{u'w'}$, where u' and w' are the gravity wave zonal and vertical velocities, respectively. (Certainly this assumption is made in all existing gravity wave drag parameterizations in climate models.) Since there is no radiation of EP flux to or from space, one can impose the upper boundary condition $\rho\overline{u'w'} \to 0$ as $z \to \infty$, in which case

$$\int_z^\infty \rho\mathcal{F}\,dz = \rho\overline{u'w'}\big|_z. \qquad (7)$$

The physical content of (7) is that any zonal momentum flux through z (the right-hand side) is entirely absorbed in the atmosphere above (the left-hand side). We shall refer to this as the ''(angular) momentum constraint.'' Substituting (7) into (6) leads to

$$\overline{w}^* = -\frac{1}{a\,\cos\phi}\frac{\partial}{\partial\phi}\left[\frac{(\cos\phi)\overline{u'w'}}{2\Omega\,\sin\phi}\right] \qquad (8)$$

(Haynes et al. 1991). Thus, the downwelling and dynamical heating induced by gravity wave drag at a given altitude z is determined entirely by the zonal momentum
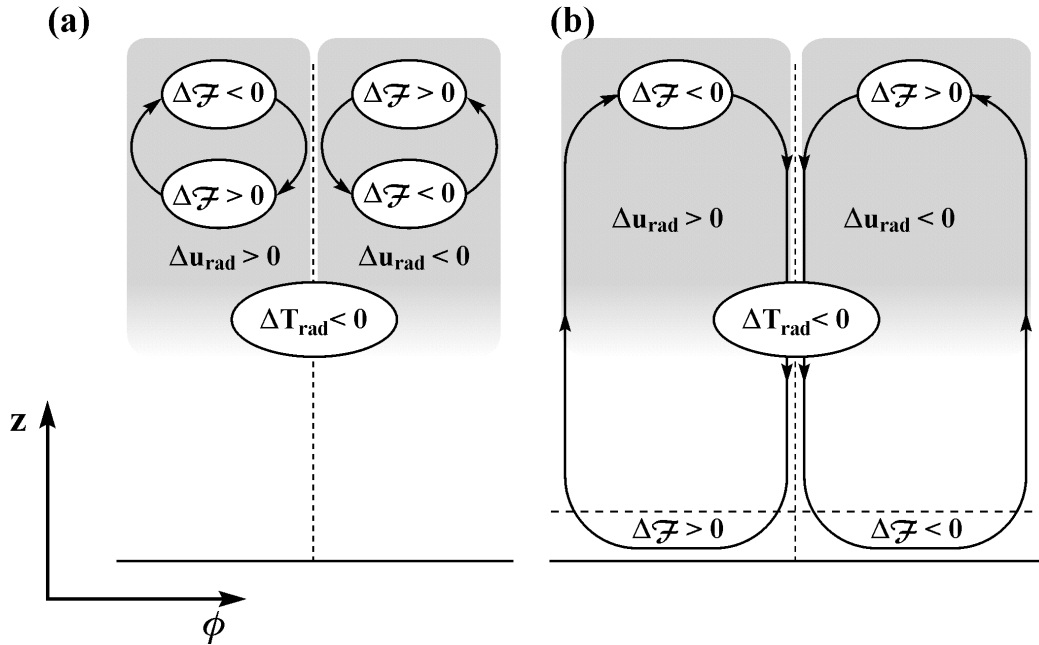
FIG. 1. Thought experiment of the steady dynamical response to a localized radiative cooling, denoted by $\Delta T_{rad} <$ 0, induced by (a) a momentum-conserving gravity wave drag parameterization and (b) a Rayleigh drag. By thermal wind, ΔT_{rad} is associated with a change in the zonal wind, Δu_{rad}, denoted by the stippled regions. The change in u_{rad} leads to a change in $\overline{u}$ that induces gravity wave drag feedbacks $\Delta \mathcal{F}$, as indicated in the figure. What is depicted are the drag *anomalies* and the associated meridional circulations. Violation of momentum conservation in (b) leads to a physically spurious downward influence. See text for detailed discussion.

flux (and its meridional gradient) passing through the same altitude z, provided this momentum flux is entirely absorbed in the atmosphere above.

The momentum constraint has important implications for possible downward influence. The current understanding of gravity wave breaking and dissipation is that, to a large extent, the momentum flux at any altitude z is determined by the atmospheric state (especially winds) at and below that altitude, not by the atmospheric state above. While this is not exactly true for nonhydrostatic gravity waves because of back reflection by turning lines, in line with most current gravity wave drag parameterizations we treat back reflection as a modification of the source spectrum and assume no sensitivity of back reflection to conditions within the middle atmosphere. Now consider the following thought experiment, depicted in Fig. 1a. A radiative perturbation is imposed, represented by a localized change in T_{rad} within the middle atmosphere—this might be from ozone depletion, volcanic aerosol loading, or solar variability. By thermal wind balance, this ΔT_{rad} is associated with a change in zonal wind, Δu_{rad}. In the Newtonian cooling approximation (and in the steady limit), the direct response of $\overline{T}$ and $\overline{u}$ to the radiative perturbation (i.e., in the absence of wave drag feedbacks) is given by ΔT_{rad} and Δu_{rad}, respectively. Depicted in Fig. 1a is a cooling, which leads to a positive wind anomaly to the south and to a negative wind anomaly to the north. These direct changes in $\overline{u}$ will generally induce changes in gravity wave drag, and

hence in $\overline{w}^*$, at and above where u_{rad} has changed (indicated by the stippled regions in Fig. 1), further modifying the zonal mean state. In the region of the positive wind anomaly, for example, waves with positive (eastward) intrinsic phase speeds will break lower down, because $|\overline{u} - c|$ is reduced, so the drag *anomaly* $\Delta \mathcal{F}$ from such waves will be a dipole with negative above and positive below, as shown on the left-hand side of Fig. 1a, with exactly cancelling torques. On the other hand, waves with negative intrinsic phase speeds will break farther up, because $|\overline{u} - c|$ is increased, so the drag anomaly from such waves will also be a dipole with negative above and positive below with exactly cancelling torques. Exactly the opposite situation obtains in the region of the negative wind anomaly, where waves with both positive and negative intrinsic phase speeds will induce dipole drag anomalies with positive above and negative below with exactly cancelling torques, as shown on the right-hand side of Fig. 1a. Thus, the nature of the drag anomalies predicted by this thought experiment is the same for waves of both positive and negative intrinsic phase speeds, which means that the argument holds for any source spectrum. Of course, with a spectrum of waves, one would obtain a distribution of such dipoles.

These drag anomalies represent the radiatively induced feedback in gravity wave drag, and lead to closed circulation cells on either side of the radiative perturbation. However, in the steady limit, Δu_{rad} is confined to altitudes above the lowest altitude, z, where ΔT_{rad} is

nonzero (Shepherd et al. 1996, Fig. 2b bottom), and therefore the circulation cells are also confined to altitudes above *z*. It follows that the gravity wave momentum flux below *z* will be unchanged. Hence, in spite of possible changes in the distribution of gravity wave drag above *z*, (8) implies that the $\overline{w}^*$, T_{dyn}, and finally $\overline{T}$ below *z* induced by gravity wave drag above *z* will be unchanged. Physically, the radiative changes above *z* may change the vertical distribution of gravity wave momentum deposition above *z* but cannot change its overall amount. This leads us to the following result.

> Remark 1: *In the time mean, radiatively induced temperature perturbations above a given altitude z cannot induce changes in zonal mean wind and temperature below z through feedbacks in gravity wave drag alone.*

Strictly speaking, remark 1 does not apply in the Tropics. However, the time mean effects of the gravity-wave-driven circulation are not considered to be important in the Tropics, for reasons discussed in section 5.

This tight constraint on $\overline{w}^*$ reflects the fact that angular momentum is conserved within the atmosphere; unlike the energy balance, which is open because of cooling to space, the angular momentum balance is closed.

The condition "through feedbacks in gravity wave drag alone" is crucial to the validity of remark 1. A change in gravity wave source strength could certainly change $\overline{w}^*$ below *z*. However, this change would have to arise from a change in atmospheric conditions at the source, and a prerequisite for such a change (in this thought experiment) is a change in $\overline{w}^*$ below *z*. Two ways in which the radiative perturbation could induce changes in the gravity wave source strength are radiative nonlocality and changes in planetary wave drag. The first mechanism is not likely to be very important for upper-stratospheric or mesospheric perturbations but could be significant for lower-stratospheric perturbations. The second mechanism is discussed in section 4.

Remark 1 has potentially important implications for climate sensitivity. It is widely accepted that gravity wave drag is required in order to obtain realistic wintertime polar stratospheric temperatures in middle-atmosphere models (Garcia and Boville 1994; Manzini and McFarlane 1998; Austin et al. 2003). Yet there is considerable debate about how to parameterize gravity wave drag, given the lack of observational constraints, and it is probably fair to say that gravity wave drag parameterizations are simply tuned to obtain a reasonable climate. One naturally worries that such tuning might adversely affect the climate sensitivity of the model. Yet remark 1 suggests that, provided the momentum constraint (7) is obeyed, the role of radiatively induced feedbacks in gravity wave drag may be considerably smaller than the role of gravity wave drag in the climate itself. Thus, the uncertainty in parameterization of gravity wave drag may not matter as much as one might have feared, with respect to climate sensitiv-

ity. (It should be emphasized that this statement applies to the parameterization of gravity wave propagation and breaking, *not* to the parameterization of gravity wave sources.)

However, parameterizations of gravity wave drag in climate models do not necessarily respect the momentum constraint of the real atmosphere. This may lead to spurious downward influence and spurious climate sensitivity. Rayleigh drag $\mathcal{F} = -\mu\overline{u}$, where μ is a damping rate, is still used by many models to represent gravity wave drag (Schnadt et al. 2002; Shine et al. 2003), even though the resulting torque is not the divergence of a flux and does not satisfy the momentum constraint. Shepherd et al. (1996, Figs. 2c, 12 top) showed how a zonal mean Rayleigh drag, either confined to an upper sponge layer or imposed throughout the atmosphere, can drive a spurious meridional circulation extending to the surface of the earth in response to a radiative perturbation. The effect is depicted in Fig. 1b. Since the zonal wind anomaly on either side of the radiative perturbation is single signed, the Rayleigh drag anomaly on either side is also single signed, which from (6) implies that there is a net vertical motion that can only close in the planetary boundary layer. In particular, there is not the cancellation within a vertical column (and associated local circulation cells) that one obtains from a momentum conserving gravity wave drag parameterization, as in Fig. 1a.

Even if a gravity wave drag parameterization is momentum conserving, it may not be implemented in a momentum-conserving way. For example, the parameterization may be imposed in the presence of a sponge layer, which absorbs much of the imposed torque and violates the momentum constraint (Shepherd et al. 1996, Fig. 3c). It is also common practice to allow any nonzero parameterized momentum flux at the model lid to escape to space. This too is unphysical and violates the momentum constraint. In order to respect the momentum constraint, the parameterized momentum flux at the model lid should be deposited within the model domain, for example, within the top few levels. While this will distort the impact of gravity wave drag near the top of the model, it will preserve the correct impact lower down, in accord with (8).

Figure 2 illustrates some of these possibilities with a quantitative calculation. It shows the zonal mean response to a switch-on radiative cooling over the pole (e.g., polar ozone depletion), in the long-term limit, calculated from the QG approximation to the time-dependent zonally symmetric TEM equations beginning from a state of rest. Thus it corresponds to the left-hand side (south of the radiative cooling) in both panels of Fig. 1. Newtonian cooling is present, with a time scale of 5 days, as is a frictional boundary layer at the earth's surface. The calculation is idealized, and the units should be regarded as arbitrary. Figure 2a shows the radiative temperature perturbation. Figure 2b shows the circulation response after 300 days in the absence of
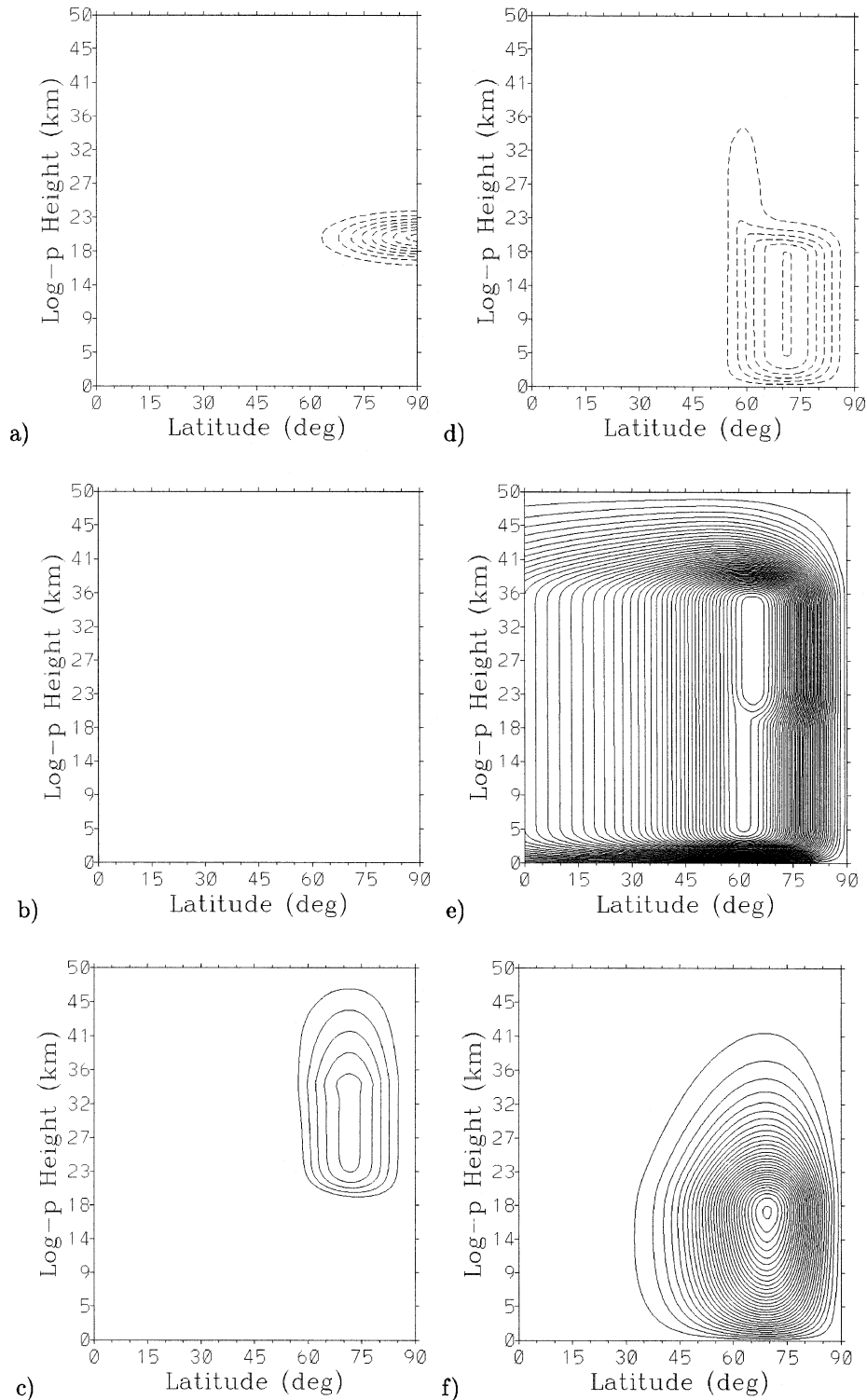
FIG. 2. Numerical simulation of the steady response to an imposed radiative cooling, calculated with a zonally symmetric QG version of the TEM equations. The calculation is global, but only the Northern Hemisphere is shown. (a) The imposed temperature change. The circulation response, with the same contour interval, for (b) no wave drag, (c) a momentum conserving implementation of the AD99 gravity wave drag parameterization, (d) a nonconservative implemention of AD99 where the momentum flux at the model lid is thrown away, (e) AD99 in the presence of a zonal mean sponge layer above 35 km, and (f) a uniform Rayleigh drag throughout the domain with a damping time scale of 100 days. Positive contours in (b)–(f) correspond to a clockwise circulation, with descent over the pole; negative contours correspond to a counterclockwise circulation. The units should be regarded as arbitrary.

any wave drag and demonstrates that radiation on its own cannot drive a steady extratropical circulation (Haynes et al. 1991). (There is actually a small transient remaining, but it is below the contour level.) Figure 2c shows the circulation response induced by the Alexander and Dunkerton (1999, hereafter AD99) gravity wave drag parameterization, with an antisymmetric source spectrum and without back reflection. The AD99 parameterization assumes that each wave deposits all of its momentum at its breaking height, where the combined lapse rate of the wave and background mean state is convectively unstable (Lindzen 1981). The breaking height is calculated using the breaking criterion

$$A \equiv \sqrt{\frac{2N}{k}} \frac{e^{(z-z_0)/2H}}{|\bar{u}-c|^{3/2}} |F_0(c)|^{1/2} = 1 \qquad (9)$$

(Alexander and Dunkerton 1999), where k is the horizontal wavenumber, N is the buoyancy frequency, H is the density scale height, c is the horizontal phase speed of the wave, z_0 is the source height, and $\rho_0 F_0(c)$ is the momentum flux at the source height (source spectrum), where ρ_0 is the density at the source height. Any nonzero momentum flux at the model lid is deposited as a vertically uniform zonal force $\rho\mathcal{F}$ in the top part of the model domain, above 35 km. In the absence of a zonal flow, the antisymmetry of the source spectrum implies that the drag is exactly zero for all z. However, the polar cooling induces an eastward zonal flow anomaly above and to the south of the cooling region, which breaks the symmetry. This forces the eastward-propagating ($c > 0$) waves to break lower down and the westward-propagating ($c < 0$) waves to break higher up. Thus, as predicted in the thought experiment, the wave drag *anomaly* is positive just above the cooling region and negative near the top of the model domain. However, the momentum constraint ensures that the drag anomalies exactly cancel in the vertical and leads to a closed meridional circulation above the cooling region, with no response below. Whether this response is quantitatively correct depends on the settings of the AD99 parameterization, but that is not the point. The important result is qualitative, namely that one obtains polar descent above the cooling region but with no downward extension of the circulation below the cooling region. Thus, Fig. 2c illustrates remark 1.

If the source spectrum were not antisymmetric, or if the initial zonal flow (before the radiative perturbation is imposed) were nonzero, then the initial state would already include nonzero gravity wave drag and a mean circulation, in accord with downward control. It is the difference between that state and the steady state obtained under the radiative perturbation that is constrained by remark 1. (In Fig. 2, the initial state has no circulation and so the circulation anomaly is the total circulation.) In such cases the details of the circulation anomaly would differ from that seen in Fig. 2c, but it

would still be confined above the level of the radiative perturbation.

Figure 2d shows the circulation response induced by the AD99 parameterization when the parameterized momentum flux at the model lid is allowed to escape to space. In this case, there is still a positive drag anomaly just above the cooling region, but the compensating negative drag anomaly is missing. As a result, the positive drag anomaly drives a circulation extending to the surface of the earth, in accord with downward control (Haynes et al. 1991). This is a spurious downward influence, a result of violating the momentum constraint. Figure 2e shows the circulation response when the AD99 parameterization is used in the presence of a zonal mean Rayleigh drag sponge layer at the top of the model. The physically correct response of Fig. 2c is evident but is augmented by a much stronger circulation driven from the sponge layer (cf. Shepherd et al. 1996, Fig. 2c top). Finally, Fig. 2f shows the circulation response from Rayleigh drag alone (in place of AD99) imposed throughout the domain. In both cases, one sees spurious downward influence. Motivated by these results, we draw our second main conclusion.

Remark 2: *In order for a gravity wave drag parameterization to respect the momentum constraint and avoid spurious downward influence, any nonzero parameterized momentum flux at the model lid must be deposited within the model domain. Moreover, there must be no zonal mean sponge layer.*

Provided the momentum constraint is respected, then it is not necessary for a model to explicitly represent the regions in which gravity wave drag occurs in order to represent the *effects* of that gravity wave drag. For example, even though mesospheric gravity wave drag is an important determinant of wintertime polar lower-stratospheric temperatures, especially in the Southern Hemisphere (Haynes et al. 1991; Garcia and Boville 1994), the time mean stratospheric downwelling induced by mesospheric gravity wave drag can be properly represented in a model with a lid at the stratopause. Furthermore, since mesospheric zonal wind and temperature changes cannot directly induce changes in the stratospheric downwelling driven by mesospheric gravity wave drag, it is not necessary to simulate mesospheric climate change (for example) in order to simulate stratospheric climate change, so long as the only coupling from the mesosphere to the stratosphere is through a gravity-wave-drag-induced circulation, and remark 2 is obeyed.

4. Downward influence through feedbacks in planetary wave drag

Remarks 1 and 2 are specific to gravity wave drag, but in fact what they assume is no meridional propagation and no reflection, assumptions that are reasonable (albeit not exact) for gravity waves. These assumptions

are definitely not applicable to planetary waves. Meridional propagation is expected from the aspect ratio of Rossby ray paths, and there is clear evidence of reflection when there is negative vertical zonal wind shear in the upper stratosphere (Perlwitz and Harnik 2003). Both mechanisms are examples of what might be called an ''active'' upper-boundary condition (one that is state dependent), as opposed to the ''passive'' upper-boundary condition assumed by the momentum constraint (with the vertical momentum flux at a level z independent of the state of the atmosphere above z).

Integrating (6) between latitude ϕ and the pole (we assume the North Pole) under the assumption $F^{(z)} \to 0$ as $z \to \infty$ (i.e., no EP flux to or from space) leads to the following constraint on steady downwelling over the polar cap at any altitude z:

$$
\int_{\phi}^{\pi/2} a\rho(\cos\phi)\overline{w}^* \, d\phi
$$

$$
= \int_{z}^{\infty} \frac{\rho \mathcal{F}}{2\Omega \, \sin\phi} \, dz
$$

$$
= -\frac{F^{(z)}}{2\Omega a \, \sin\phi} + \frac{1}{\Omega a^2 \, \sin(2\phi)} \frac{\partial}{\partial \phi}\left[\int_{z}^{\infty} (\cos\phi)F^{(\phi)} \, dz\right],
$$

$$(10)$$

where $F^{(z)}$ and $F^{(\phi)}$ are respectively the vertical and meridional components of the EP flux. This may be regarded as a form of the momentum constraint relevant to planetary waves, although it applies in general. The relation (10) states that the downwelling is controlled by the vertical EP flux at z and ϕ, as well as by the meridional convergence of area-weighted EP flux at ϕ and all altitudes above z. The latter is proportional to the wave-induced meridional mass flux into or out of the polar cap. Reflection can affect $F^{(z)}$; meridional propagation can affect $F^{(\phi)}$. To the extent that these quantities *did not* change under a radiative perturbation, then the momentum constraint would hold and remarks 1 and 2 would apply. In such a case, once again downward control would not mean downward influence, although in a model spurious downward influence could arise if any nonzero EP flux at the top of the model was somehow discarded (as might be possible in a two-dimensional model) or damped in a zonal mean sponge layer. This then leads to our final result.

> Remark 3: *In the time mean, true downward influence from a radiative perturbation can only arise through changes in reflection and meridional propagation properties of planetary waves.*

The changes in planetary wave properties referred to in remark 3 include both the direct response to the radiative perturbation, and the indirect response to the gravity wave drag changes induced by that perturbation. If the assumption of no gravity wave reflection or meridional propagation is relaxed, then remark 3 continues to hold but now includes gravity waves as well as planetary waves. In the real atmosphere, it is expected that the potential for true downward influence is far greater for planetary waves than it is for gravity waves.

5. Summary and discussion

In the time mean, downward control (Haynes et al. 1991) implies that the vertical mass flux and dynamical heating at any altitude z depends on the wave drag (and its meridional gradient) above that level. In the special case of gravity wave drag with no meridional propagation, and assuming no exchange of angular momentum with space, the wave drag above z depends solely on the zonal momentum flux at z. This constraint, which is referred to as the ''(angular) momentum constraint,'' is mathematically represented in relation (7) and reflects the fundamental conservation law for angular momentum in a viscous fluid. It implies that the vertical mass flux and dynamical heating at z induced by gravity wave drag above z is determined by the gravity wave momentum flux at z (Haynes et al. 1991).

Under the further assumption that the gravity wave momentum flux at z depends only on the atmospheric conditions at and below z, and is independent of the conditions above, the momentum constraint provides a strong limitation on gravity wave drag feedbacks to radiative perturbations. In particular, radiatively induced temperature perturbations above a given altitude z cannot induce time mean changes in zonal mean wind and temperature below z through feedbacks in gravity wave drag alone (remark 1). This prediction of exactly zero downward influence—in spite of downward control— is confirmed by a numerical calculation with a state-of-the-art gravity wave drag parameterization (Fig. 2c). Thus, despite the many uncertainties in the parameterization of gravity wave drag, the role of gravity wave drag in climate perturbations may be much more limited than its role in climate itself.

Remark 1 limits the possibilities for downward influence from the mesosphere. In order for a gravity wave drag parameterization in a model to respect the momentum constraint and avoid spurious downward influence, any nonzero parameterized momentum flux at the model lid must be deposited within the model domain, and there must be no zonal mean sponge layer (remark 2). Figures 2d–f show how violation of these conditions leads to spurious downward influence. For a zonal mean (two-dimensional) model formulated in terms of the TEM equations, an alternative strategy would be to replace the usual upper-boundary condition $\rho\overline{w}^* = 0$ with (8).

Remark 1 means that it is not necessary to simulate the region in which gravity wave drag occurs in order to represent the effects of that drag on regions below, provided the momentum constraint is respected. Thus, for example, it would not be necessary to simulate me-

sospheric climate change in order to simulate stratospheric climate change, if the only coupling from the mesosphere to the stratosphere was through the gravity-wave-drag-induced circulation, and remark 2 was obeyed.

For planetary waves, a weaker form (10) of the momentum constraint exists that does not prohibit downward influence but limits the mechanisms by which it can occur: in the time mean, downward influence from a radiative perturbation can only arise through changes in reflection and meridional propagation properties of planetary waves (remark 3). This suggests that analysis of climate change simulations should focus on those aspects of planetary wave behavior. However, it is equally (if not more) important for such analysis to also assess whether the momentum constraint is otherwise respected in the model, in order to avoid the possibility of spurious downward influence in the simulations.

The analysis in this paper makes a number of assumptions that are not exactly true but are believed to be reasonable for understanding the climate of the middle atmosphere. First, linearity of the wave-driven circulation is assumed; in this case, the time mean state obeys the steady TEM equations. This assumption is quite reasonable in the extratropics, where angular momentum contours depart only slightly from solid-body rotation (Haynes et al. 1991), but becomes problematic in the Tropics. However the time mean effects of the gravity-wave-driven circulation are only considered to be of importance in the extratropics; in the Tropics, wave drag goes mainly into zonal flow acceleration rather than into meridional circulation, and the response is inherently transient (Andrews et al. 1987). The remarks in this paper do not apply to the quasi-biennial oscillation, for example, although the quasi-biennial oscillation is in any case not usually regarded as a mechanism for downward influence because according to our current understanding the downward propagation of zonal wind anomalies is only propagation of phase, not of information (Andrews et al. 1987). (The possibility that the quasi-biennial oscillation could induce changes in extratropical stratospheric planetary wave drag with a surface response falls within the sort of downward influence discussed in section 4.)

Second, thermal radiative locality is assumed, here in the strict form of Newtonian cooling, which leads to (2). This "cooling to space" approximation is widely used in middle-atmosphere applications (Liou 2002; Andrews et al. 1987). In practice, thermal radiative nonlocality means that a nonzero $\overline{w}^*$ at a given z can lead to a dynamical temperature response below z. However, downward nonlocality for a well-mixed emitter like CO_2 will be limited by density, and we would expect it to be restricted to a scale height or so. The strength of the radiative internal heat exchange of a given level with the atmosphere above can be estimated from the values of parameterization coefficients of a matrix parameterization for the 15-μm CO_2 band cooling (Fomichev et

al. 1998). As seen from Tables 1–8 of Fomichev et al. (1998), parameterization coefficients describing radiative heat exchange with levels located more than a scale height above the level in question are at least an order of magnitude smaller than the coefficients describing the cooling-to-space term.

Third, following Haynes et al. (1991), the only relaxational momentum source or sink is assumed to be in the planetary boundary layer. In the thermosphere, molecular viscosity becomes important. However, because of the anisotropic aspect ratio of atmospheric disturbances, vertical diffusion should greatly dominate over horizontal diffusion, and in that case there is no net momentum change within each column because vertical diffusion obeys the momentum constraint. Thus, molecular diffusion does not alter the present results. On the other hand, ion drag is a relaxational process so far as the neutral atmosphere is concerned, and would violate the momentum constraint. Ion drag plays a significant role in the momentum balance above about 130 km (Fomichev et al. 2002), but its effects are likely to be quite negligible in the middle atmosphere.

Probably the most significant assumptions made here are those of no meridional propagation and no sensitivity of back reflection to conditions within the middle atmosphere. In the present state of knowledge these are not believed to be very significant effects, and indeed they are not even represented in most current gravity wave drag parameterizations. However, one implication of the present analysis is that any true downward influence in climate sensitivity from gravity wave drag alone must come from these mechanisms, so they may warrant closer investigation.

REFERENCES

Alexander, M. J., and T. J. Dunkerton, 1999: A spectral parameterization of mean-flow forcing due to breaking gravity waves. *J. Atmos. Sci.,* **56,** 4167–4182.

Andrews, D. G., J. R. Holton, and C. B. Leovy, 1987: *Middle Atmosphere Dynamics.* Academic Press, 489 pp.

Austin, J., and Coauthors, 2003: Uncertainties and assessments of chemistry-climate models of the stratosphere. *Atmos. Chem. Phys.,* **3,** 1–27.

Beagley, S. R., J. de Grandpré, J. N. Koshyk, N. A. McFarlane, and T. G. Shepherd, 1997: Radiative–dynamical climatology of the first-generation Canadian Middle Atmosphere Model. *Atmos–Ocean,* **35,** 293–331.

Boville, B. A., 1995: Middle atmosphere version of CCM2

(MACCM2): Annual cycle and interannual variability. *J. Geophys. Res.,* **100,** 9017–9039.

Codoban, S., and T. G. Shepherd, 2003: Energetics of a symmetric circulation including momentum constraints. *J. Atmos. Sci.,* **60,** 2019–2028.

Fomichev, V. I., J.-P. Blanchet, and D. S. Turner, 1998: Matrix parameterization of the 15 μm CO_2 band cooling in the middle and upper atmosphere for variable CO_2 concentration. *J. Geophys. Res.,* **103,** 11 505–11 528.

——, W. E. Ward, S. R. Beagley, C. McLandress, J. C. McConnell, N. A. McFarlane, and T. G. Shepherd, 2002: The extended Canadian Middle Atmosphere Model: Zonal-mean climatology and physical parameterizations. *J. Geophys. Res.,* **107,** 4087, doi:10.1029/2001JD000479.

Fritts, D. C., and M. J. Alexander, 2003: Gravity wave dynamics and effects in the middle atmosphere. *Rev. Geophys.,* **41,** 1003, doi:10.1029/2001RG000106.

Garcia, R. R., 1987: On the mean meridional circulation of the middle atmosphere. *J. Atmos. Sci.,* **44,** 3599–3609.

——, and B. A. Boville, 1994: "Downward control" of the mean meridional circulation and temperature distribution of the polar winter stratosphere. *J. Atmos. Sci.,* **51,** 2238–2245.

Haynes, P. H., C. J. Marks, M. E. McIntyre, T. G. Shepherd, and K. P. Shine, 1991: On the "downward control" of extratropical diabatic circulations by eddy-induced mean zonal forces. *J. Atmos. Sci.,* **48,** 651–678.

Holton, J. R., P. H. Haynes, M. E. McIntyre, A. R. Douglass, R. B. Rood, and L. Pfister, 1995: Stratosphere–troposphere exchange. *Rev. Geophys.,* **33,** 403–439.

Lindzen, R. S., 1981: Turbulence and stress owing to gravity wave and tidal breakdown. *J. Geophys. Res.,* **86,** 9707–9714.

Liou, K.-N., 2002: *An Introduction to Atmospheric Radiation.* 2d ed. Academic Press, 583 pp.

Manzini, E., and N. A. McFarlane, 1998: The effect of varying the source spectrum of a gravity wave parameterization in a middle atmosphere general circulation model. *J. Geophys. Res.,* **103,** 31 523–31 539.

McLandress, C., 1998: On the importance of gravity waves in the middle atmosphere and their parameterization in general circulation models. *J. Atmos. Sol.-Terr. Phys.,* **60,** 1357–1383.

Perlwitz, J., and N. Harnik, 2003: Observational evidence of a stratospheric influence on the troposphere by planetary wave reflection. *J. Climate,* **16,** 3011–3026.

Schnadt, C., M. Dameris, M. Ponater, R. Hein, V. Grewe, and B. Steil, 2002: Interaction of atmospheric chemistry and climate and its impact on stratospheric ozone. *Climate Dyn.,* **18,** 501–517.

Semeniuk, K., and T. G. Shepherd, 2001: Mechanisms for tropical upwelling in the stratosphere. *J. Atmos. Sci.,* **58,** 3097–3115.

Shepherd, T. G., 2000: The middle atmosphere. *J. Atmos. Sol.-Terr. Phys.,* **62,** 1587–1601.

——, 2003: Hamiltonian dynamics. *Encyclopedia of Atmospheric Sciences,* J. R. Holton et al., Eds., Academic Press, 929–938.

——, K. Semeniuk, and J. N. Koshyk, 1996: Sponge layer feedbacks in middle-atmosphere models. *J. Geophys. Res.,* **101,** 23 447–23 464.

Shine, K. P., 1989: Sources and sinks of zonal momentum in the middle atmosphere diagnosed using the diabatic circulation. *Quart. J. Roy. Meteor. Soc.,* **115,** 265–292.

——, and Coauthors, 2003: A comparison of model-simulated trends in stratospheric temperatures. *Quart. J. Roy. Meteor. Soc.,* **129,** 1565–1588.

Snieder, R. K., and S. B. Fels, 1988: The flywheel effect in the middle atmosphere. *J. Atmos. Sci.,* **45,** 3996–4004.