TECHNICAL MEMORANDUM

593

The representation of non-orographic gravity waves in the IFS Part II: A physically based spectral parametrization

Andrew Orr

Research Department

May 2009

ECMWF
European Centre for Medium-Range Weather Forecasts
Europäisches Zentrum für mittelfristige Wettervorhersage
Centre européen pour les prévisions météorologiques à moyen terme

Series: ECMWF Technical Memoranda

A full list of ECMWF Publications can be found on our web site under:
http://www.ecmwf.int/publications/

Contact: library@ecmwf.int

European Centre for Medium Range Weather Forecasts
Shinfield Park, Reading, Berkshire RG2 9AX, England

Abstract

The middle atmosphere climate of cycle 33R1 of the IFS is simulated using the Scinocca (2003) parametrization of the momentum deposition from non-orographic gravity waves (hereafter SO3). A comparison is made with a basic cycle 33R1 simulation using Rayleigh friction, which is the standard method of parametrizing the effects of non-orographic gravity waves in this cycle of the IFS (hereafter RF). The SO3 scheme assumes hydrostatic and non-rotational wave dynamics, and describes the vertical evolution of a broad, constant, isotropic spectrum of gravity waves, emanating from the troposphere. The scheme models critical level filtering, conservative propagation, and nonlinear dissipation, which is dealt with in an empirical fashion by limiting the growth of the gravity wave spectrum so as not to exceed some saturated spectrum proportional to m^{-3}, where m is the vertical wavenumber. The SO3 simulation shows a number of improvements to the zonal-mean circulation relative to the RF simulation; noticeably an improved overturning circulation resulting in more realistic winter westerly winds, a reduction in the southern winter polar stratosphere cold bias, and an improvement in the capture of summer easterlies in the stratosphere and mesosphere. Estimates of gravity wave drag and gravity wave momentum flux agree well with the limited observations available. Investigation into the response of stationary planetary waves indicates a good agreement between SO3 simulated and observed amplitudes and an improved representation of planetary wave driving of the zonal-mean flow. However, although the capture of the tropical circulation is somewhat improved compared to the RF simulation, the SO3 simulation is marked by a failure to capture the westerly components of the quasi-biennial and semi-annual oscillations (QBO and SAO), resulting in a strong easterly bias here. There was a failure to simulate a QBO-like oscillation with just resolved forcing, which suggested the need for increased vertical resolution in the tropical stratosphere and mesosphere to properly resolve equatorially-trapped Kelvin and Rossby waves. A review of available observations show a strong correlation between gravity waves and (deep) convection in the tropics, suggesting that the SO3 scheme may benefit from a latitudinally varying source spectrum. An additional comparison was made with a cycle 33R1 simulation with the Doppler spread parametrization of gravity waves replacing Rayleigh friction. This scheme controls momentum deposition by Doppler shifting due to the background winds and from wave-induced fluctuations of the gravity wave winds themselves. Despite showing an improved representation of the QBO, this scheme produced unrealistically strong values of gravity wave drag, resulting in an overly strong meridional circulation and excessive downwelling, and an overly warm and elevated winter stratopause.

1. Introduction

Prior to and including cycle 33R1 of the ECMWF Integrated Forecasting System (IFS), the model uses Rayleigh friction and shows serious systematic biases in middle atmosphere climate simulations. First, there is a tendency for the circulation in the extratropics to be unrealistically close to radiative equilibrium. This problem is most apparent where planetary wave driving of the stratosphere is weak, i.e. during southern winter and early northern winter, and is characterised by simulated winter polar temperatures in the stratosphere being much too cold and the winter westerly jet much too strong (Shine, 1989; Garcia and Boville, 1994). This is due to the absence of small scale non-orographic gravity waves at climate-like resolution, which are generally unresolved or under-resolved by the IFS and other general circulation models (GCMs). Consequently, the model eddy activity is overly weak and unable to drive the required overturning circulation, resulting in insufficient dynamical warming of the winter pole (Hamilton et al., 1995, 1999). This problem is partially alleviated by the inclusion of some additional drag on the zonal-mean flow, U, in the form of Rayleigh friction above the stratopause, such that

$$F_u = -K(z)U \tag{1}$$

which is formulated as a drag force proportional to U, where the damping rate, $K(z)$, increases with height, z, and F_u is the resulting drag or frictional forcing which is added to the right-hand-side of the zonal momentum equation (e.g. McLandress, 1998). However, the large extratropical middle atmosphere biases detailed in a companion memorandum by Orr and Wedi (2009) suggest that using Rayleigh friction is unsatisfactory. Second, the interannual variability of the tropical stratospheric circulation is not captured, most notably an absence of the quasi-biennial and semi-annual oscillations (QBO and SAO, respectively). This is a consequence of the failure of Rayleigh friction to supply the required westerly momentum flux. These two issues suggest that the extra drag should take the form of a more physically based parametrization of the anticipated subgrid-scale non-orographic gravity wave effects.

Non-orographic gravity waves (which hereafter are referred to as 'gravity waves') can have vertical wavelengths, λ_z, which vary from less than one to tens of kilometres, and horizontal wavelengths, λ_l, which vary from tens to thousands of kilometres. The term non-orographic implies that the waves are non-stationary, and so induce waves with non-zero phase speeds. In the middle atmosphere, the vertical wavenumber, $m\left(=2\pi/\lambda_z\right)$, spectrum of horizontal winds induced by gravity waves is proportional to m^{-t}, with $t=3$ (VanZandt, 1982) and $m > m_*$ (i.e. large-m), where m_* is a transitional wavenumber estimated by Allen and Vincent (1995) to correspond to $\lambda_z \approx 2-3$ km in the troposphere and stratosphere. This so-called 'tail spectrum' is largely independent of both time and location (VanZandt, 1982). By contrast, gravity wave variations in both time and location are observed in the troposphere, which is where the gravity waves (mostly) originate. Here they are generated by mechanisms such as convection, frontogenesis, and jet stream activity (e.g. Fritts and Nastrom, 1992; Eckermann and Vincent, 1993; Ratnam et al., 2004; Jiang et al., 2005, Lane et al., 2008), and consist of a spectrum of horizontal winds proportional to m^s, with $s=1$ (Allen and Vincent, 1995) and $m < m_*$ (i.e. small-m). However, the value of $s=1$ is not well defined by observations, as these gravity waves have small amplitudes and large λ_z, and are thus difficult to measure.

It is now common to suggest that the apparent universal shape of the vertical wavenumber spectrum of horizontal wind variance at large-m is a result of 'saturation' (Dewan and Good, 1986). The decrease of air density, ρ, with height means that as the gravity waves propagate vertically their amplitude increases. At mesospheric levels, the waves can reach such large amplitudes that they exceed the amount consistent with the internal wave system remaining stable, i.e. the wave becomes saturated (Smith et al., 1987). With further ascent the wave becomes unstable, resulting in nonlinear breakdown and turbulent dissipation (which hereafter will be referred to as nonlinear dissipation). This acts to reduce the amplitude of the wave until it is again stable, resulting in the wave amplitude remaining constant above the breaking level and the spectrum of horizontal wind variance characterised by a m^{-3} tail at large-m. The deposited momentum induces a force or drag, which drags the mean flow toward the dissipating waves phase speed, modifying the general circulation of the middle atmosphere (Holton, 1982, 1983, 2004; Garcia and Solomon, 1985). Garcia and Boville (1994) showed that through adiabatic expansion, the gravity wave drag induced downwelling can affect winter polar temperatures down as far as the lower stratosphere. In addition, the breaking waves generate turbulence, which acts as a diffusion to further modify the mean circulation, and mixing.

As stated earlier, gravity waves are also important in driving the equatorial QBO and SAO. Using satellite observations, Hitchman and Leovy (1986) suggested that large-scale planetary waves only contributed 20 to 70% of the westerly momentum flux required to produce these oscillations, and that the source of the remaining momentum flux must come from a broad spectrum of gravity waves (Lindzen and Holton, 1968; Alexander and Holton, 1997). The gravity waves break near the critical levels of the alternating descending eastward and westward shear zones of the QBO. Waves which are unfiltered continue to propagate upwards and can break near the critical levels of the SAO.

Progress in developing a more physically based gravity wave parametrization has been hampered by limited observations and uncertain theoretical understanding, which is why for many years Rayleigh friction has been used. However, considerable advances have been made recently in both field observations and theoretical understanding, leading to a number of such physically realistic schemes (e.g. Medvedev and Klaassen, 1995; Hines 1997a, b; Alexander and Dunkerton, 1999; Warner and McIntyre, 1996, 2001; Scinocca, 2002, 2003). These launch a spectrum of upwardly propagating gravity waves which enter the lower stratosphere (which hereafter is referred to as the 'launch spectrum'), consisting largely of waves which are not resolved. Typically, the launch spectrum is considered to be constant, as the available observations, despite showing space and time variations, are still too limited to adequately constrain it. This simplifies the problem so that only the vertical propagation of the wave field and its eventual dissipation and resulting drag is parametrized. Studies such as Manzini and McFarlane (1998), Medvedev et al. (1998), Scaife et al. (2002), Formichev et al. (2002), Giorgetta et al. (2002), and Scinocca (2002, 2003), have all demonstrated an alleviation of middle atmosphere biases in GCMs through implementation of a physically based parametrization. For a review of the effects of gravity waves, see Fritts and Alexander (2003). For a review of their parametrization, see McLandress (1998).

The Warner and McIntyre (1996) scheme uses non-hydrostatic and rotational wave dynamics, and includes a carefully considered framework which separates 'well-understood' aspects of the parametrization problem from aspects which are 'ill-understood'. The clearly understood aspects are conservative propagation of gravity waves, critical level filtering, and total internal reflection. The ill-understood aspect is the deposition of momentum arising from the nonlinear dissipation of the gravity waves, which is dealt with in an empirical manner. This is discussed further in the next section. However, the Warner and McIntyre (1996) scheme is considered too expensive to implement in an operational model. To address this issue, Scinocca (2003) developed a simplified version of the scheme by assuming hydrostatic and non-rotational wave dynamics. This efficient implementation is suitably fast for operational use, and was shown to alleviate much of the middle atmosphere wind biases that occurred in the Canadian Middle Atmosphere model implementing the Hines (1997a, b) Doppler spread parametrization of gravity waves (Scinocca, 2002, 2003). This memorandum examines the impact on the IFS middle atmosphere climate when Rayleigh friction is replaced by gravity wave forcing represented by this scheme.

The outline of this memorandum is as follows: Section 2 describes the Scinocca (2003) scheme (hereafter SO3). Section 3 details the experimental setup and the observations used for validation. Two 8-year simulations were performed, one with Rayleigh friction (hereafter RF), and another with the SO3 scheme in place of Rayleigh friction. Section 4 examines the response of the circulation in the middle atmosphere by firstly looking at changes to the mean circulation and temperature structure. Changes to the average residual meridional and residual vertical velocity are also analysed, where the residual circulation is the Eulerian

circulation minus the contribution from model resolved eddy heat fluxes (Andrews et al., 1987). The residual mean vertical motion therefore approximately represents the diabatic circulation in the meridional plane, and the residual mean meridional circulation approximates the mean motion of air parcels, and unlike the conventional Eulerian mean, provides an approximation to the mean advective transport (Andrews et al., 1987; Holton, 2004). To examine the stationary behaviour of the atmosphere, changes to the amplitude of large-scale stationary planetary waves (i.e. waves that do not change in time and whose zonal average is zero) and their associated Eliassen-Palm (EP) flux divergence are computed in section 5, where the EP-flux divergence quantifies the acceleration/deceleration of the waves on the zonal mean circulation (Edmon et al., 1980). Section 6 details the representation of the QBO and SAO. A comparison of simulated and absolute values of gravity wave momentum flux derived from satellite data is detailed in section 7. Section 8 reports on the sensitivity of the middle atmosphere circulation to varying the gravity wave launch spectrum. Section 9 details the results of a third 8-year simulation in which gravity wave forcing is represented by the Hines (1997a, b) Doppler spread parametrization scheme. The conclusions of this study are presented in section 10.

2. The SO3 gravity wave scheme

2.1 The effect of hydrostatic non-rotational wave dynamics

As stated, the SO3 gravity wave drag scheme employs simplified hydrostatic non-rotational wave dynamics. Before fully presenting this framework, it is useful to have some understanding of the effects of neglecting non-hydrostatic and rotational wave dynamics. This issue was examined by Scinocca (2002).

The dispersion relation for an individual non-hydrostatic gravity wave in the presence of rotation in a stratified atmosphere is (Gill, 1982)

$$m^2 = \frac{k^2 N^2}{\tilde{\omega}^2} \frac{(1-\tilde{\omega}^2/N^2)}{(1-f^2/\tilde{\omega}^2)} \qquad (2)$$

where N is the buoyancy frequency, $k\left(=2\pi/\lambda_l\right)$ is the horizontal wavenumber, f is the inertial frequency, and $\tilde{\omega} = \omega - kU$ is the intrinsic frequency (with ω the ground based frequency, and U now defined as the background wind speed in the direction of wave propagation). From (2) it can be seen that wavelike solutions (i.e. $m^2 > 0$) are defined for intrinsic frequencies that fall in the range $f^2 < \tilde{\omega}^2 < N^2$, and that critical levels occur when the intrinsic frequency approaches the inertial frequency. Moreover, non-hydrostatic dynamics allows the possibility of some portions of the gravity wave spectrum becoming evanescent as the launch spectrum encounters varying winds and buoyancy frequencies as it propagates up into the middle atmosphere. Such portions of the wave spectrum undergo internal reflection and are reflected back downwards, and are therefore unable to interact with the flow at upper levels. Warner and McIntyre (1996) referred to this process as 'back-reflection'. This was shown by Scinocca (2002) to reduce the net value of momentum flux at the launch altitude by as much as 75% by removing those portions of the wave spectrum which are back-reflected. Furthermore, as the portions of the wave spectrum removed are dependent on latitudinal and seasonal variation of winds and buoyancy frequencies, the resulting 'effective' launch spectrum similarly displays systematic latitudinal and seasonal variations even though the initial launch spectrum was independent of time and geographic location.

By contrast, the simplified dispersion relation for hydrostatic gravity waves in the absence of rotation is

$$m^2 = \frac{k^2 N^2}{\tilde{\omega}^2} = \frac{N^2}{\tilde{c}^2} \tag{3}$$

where $\tilde{c} = c - U$ is the intrinsic phase speed of the gravity wave, and c is the ground based phase speed. This is recovered by substituting into (2) the limits of

$$\tilde{\omega}^2 / N^2 \to 0 \quad \text{(hydrostatic wave dynamics)} \tag{4}$$

and

$$f^2 / \tilde{\omega}^2 \to 0 \quad \text{(no rotation)} \tag{5}$$

From (3) it is apparent that wavelike solutions are defined for intrinsic frequencies that fall in the range $0 < \tilde{\omega}^2 < \infty$, and that that critical levels occur when the intrinsic frequency approaches zero. Moreover, internal reflection is not possible with hydrostatic dynamics, i.e. the amount of momentum flux available to the flow above is identical to the amount launched, and there is no seasonal or latitudinal variation of the launch spectrum (Scinocca, 2002).

Despite this, Scinocca (2002) showed that realistic results using a *hydrostatic* version of the Warner and McIntyre (1996) parametrization could be achieved by simply reducing the amount of launch momentum flux, suggesting that the reduction in the net momentum flux is more important than the inclusion of latitudinal and seasonal variations. Scinocca (2002) further showed that results were relatively insensitive to the inclusion of rotation. It was these encouraging findings which led to development of the hydrostatic non-rotational SO3 scheme. The remainder of this section details the derivation of this scheme.

2.2 Launch spectrum and conservative propagation

The initial launch spectrum is assumed to be azimuthally isotropic (i.e. the net input momentum flux at the launch level is zero) and globally uniform. In any azimuth, ϕ, the launch spectrum is specified by the total wave energy per unit mass, the spectral density of which is assumed to be of the generalised Desaubies form (Fritts and VanZandt, 1993):

$$\tilde{E}(m,\tilde{\omega},\phi) = B\left(\frac{m}{m_*}\right)^s \frac{N^2 \tilde{\omega}^{-p}}{1-\left(\frac{m}{m_*}\right)^{s+t}} \tag{6}$$

Where B and p are constants, and $\tilde{E}$ is a function of $(m,\tilde{\omega},\phi)$, referred to as $m-\tilde{\omega}$ space. The above expression is separable in terms of both m and $\tilde{\omega}$. The dependence on m is $\tilde{E} \propto m^s$ for $m < m_*$ and $\tilde{E} \propto m^{-t}$ for $m > m_*$. The dependence on $\tilde{\omega}$ is $\hat{E} \propto \tilde{\omega}^{-p}$ for all $\tilde{\omega}$, with $1 \le p \le 5/3$ (Fritts and VanZandt, 1993; Warner and McIntyre, 1996).

Although $\tilde{E}(m,\tilde{\omega},\phi)$ specifies the form of the launch spectrum, it is the momentum flux spectral density, $\rho\tilde{F}(m,\tilde{\omega},\phi)$, which is required. This is obtained from (6) by the group velocity rule:

$$\rho\tilde{F}(m,\tilde{\omega},\phi) = \rho c_{gz}\frac{k}{\tilde{\omega}}\tilde{E}(m,\tilde{\omega},\phi) \tag{7}$$

where $c_{gz} = \partial\tilde{\omega}/\partial m$ is the vertical group velocity, with $c_{gz} = \tilde{\omega}/m$ for hydrostatic dynamics.

However, m and $\tilde{\omega}$ are not invariant as they change as the launch spectrum propagates vertically through the height-varying background wind, $U(z)$, and buoyancy frequency, $N(z)$. Therefore, spectral elements in $m-\tilde{\omega}$ space $\left(dmd\tilde{\omega}d\phi\right)$ are also not invariant to vertical changes in $U(z),N(z)$. Consequently, the densities $\tilde{E}(m,\tilde{\omega},\phi)$ and $\rho\tilde{F}(m,\tilde{\omega},\phi)$ are not conserved for conservative propagation (i.e. in the absence of dissipative processes), and therefore the $m-\tilde{\omega}$ coordinate framework is unsuitable (Warner and McIntyre, 1996). By contrast, assuming that the wind and buoyancy frequency are taken to be horizontally uniform and time-independent, ω and k are invariant to vertical changes of $U(z),N(z)$. Therefore spectral elements in $\omega-k$ space $\left(d\omega dkd\phi\right)$ are also invariant. Consequently, the density $\rho F(k,\omega,\phi)$ is conserved for conservative propagation. Moreover, the relation $c=\omega/k$ [which is derived from (3)] allows the density to be a function only of the conserved variable c, making this the preferred coordinate framework in which to base the scheme. The new expression for the momentum flux density as a function of c is obtained by firstly transforming $\rho\tilde{F}(m,\tilde{\omega},\phi)$ from $m-\tilde{\omega}$ to $c-\tilde{\omega}$ space, and secondly by integrating out the $\tilde{\omega}$ dependence.

In the first step, the transformation from $m-\tilde{\omega}$ to $c-\tilde{\omega}$ space is via

$$\rho F(c,\tilde{\omega},\phi) = J\rho\tilde{F}(m,\tilde{\omega},\phi) \tag{8}$$

where the Jacobian, J, is given by

$$J = \frac{\partial\left(m,\tilde{\omega},\phi\right)}{\partial\left(c,\tilde{\omega},\phi\right)} = \frac{m^2}{N} \tag{9}$$

Combining (6), (7), (8), and (9), gives

$$\rho F(c,\tilde{\omega},\phi) = \rho B\left(\frac{m}{m_*}\right)^s \frac{m^2}{1+\left(\frac{m}{m_*}\right)^{s+t}}\tilde{\omega}^{1-p} \tag{10}$$

The second step integrates out the $\tilde{\omega}$ dependence of (10). Although hydrostatic dynamics implies $m^2>0$ (i.e. wavelike disturbances) for all $\tilde{\omega}\neq 0$, the intrinsic frequency at launch level, z_0, is restricted to the more physically realistic range of $f\le\tilde{\omega}\le N_0$, consistent with non-hydrostatic rotational dynamics.

(Hereafter the subscript '0' refers to values of parameters at launch level, $z = z_0$; the subscript '1' refers to values of parameters at the level $z = z_1$; etc.) Thus f and N_0 provide the appropriate lower and upper bounds for the integration at this level. However, for $z > z_0$, it is possible for $\tilde{\omega}$ to be Doppler shifted outside $f \le \tilde{\omega} \le N$, such that $0 \le \tilde{\omega} \le \infty$. Thus 0 and ∞ provide the appropriate lower and upper bounds for the integration at this level. This is equivalent to integrating (10) over the range $f \le \tilde{\omega} \le N_0$ on the launch level and $\tilde{\omega}_{low}(z) \le \tilde{\omega} \le \tilde{\omega}_{hi}(z)$ on subsequent levels, with the resulting expression given as:

$$\rho\overline{F}(c,\phi) = \rho BI(z;p)\left(\frac{m}{m_*}\right)\frac{m^2}{1+\left(\frac{m}{m_*}\right)^{s+t}} \tag{11}$$

where the overbar denotes integration with respect to $\tilde{\omega}$, and

$$I(z;p) = \int_{\tilde{\omega}_{low}(z)}^{\tilde{\omega}_{hi}(z)} \tilde{\omega}^{1-p} d\tilde{\omega} = \frac{\tilde{\omega}_{hi}(z)^{2-p} - \tilde{\omega}_{low}(z)^{2-p}}{2-p} \tag{12}$$

Thus for given profiles of $U(z), N(z)$, the scheme requires the determination of values of $\tilde{\omega}_{low}(z)$ and $\tilde{\omega}_{hi}(z)$. Using $c = \omega / k$, the intrinsic frequency at the launch level can be denoted as $\tilde{\omega}(z_0) = k(c - U_0)$. At higher elevations, the same spectral element will have the intrinsic frequency $\tilde{\omega}(z) = k(c - U(z))$, which implies the relationship

$$\tilde{\omega}(z) = \tilde{\omega}(z_0)\left(\frac{c-U}{c-U_0}\right) \tag{13}$$

which allows (12) to be re-written as

$$I(z;p) = \left(\frac{c-U}{c-U_0}\right)^{2-p} \frac{N_0^{2-p} - f^{2-p}}{2-p} \tag{14}$$

Finally, performing the Galilean transformation

$$\hat{U} = U - U_0,\ \hat{c} = c - U_0 \tag{15}$$

and taking $t = 3$ and using (3) to eliminate *m*, allows (11) to be expressed in terms of the conserved variable $\hat{c}$

$$\rho\overline{F}(\hat{c},\phi) = \rho A \frac{\hat{c}-\hat{U}}{N}\left(\frac{\hat{c}-\hat{U}}{\hat{c}}\right)^{2-p} \frac{1}{1+\left(\frac{m_*(\hat{c}-\hat{U})}{N}\right)^{s+3}} \tag{16}$$

where the constant A is all terms which are independent of height, and is given by

$$A = Bm_*^3 \frac{N_0^{2-p} - f^{2-p}}{2-p} \tag{17}$$

$\rho\bar{\bar{F}}(\hat{c},\phi)$ is discretized using n_ϕ azimuths equally spaced around the azimuth circle (i.e. the source spectrum is isotropic), and $n_{\hat{c}}$ horizontal phase speeds $\hat{c}$. To reduce the value of $n_{\hat{c}}$, a coordinate stretch is used to increase the resolution at large $\hat{c}$ (i.e. small-*m*), which is the portion of the wave spectrum which is most crucial for gravity wave drag in the mesosphere. This is discussed further in section 2.5.

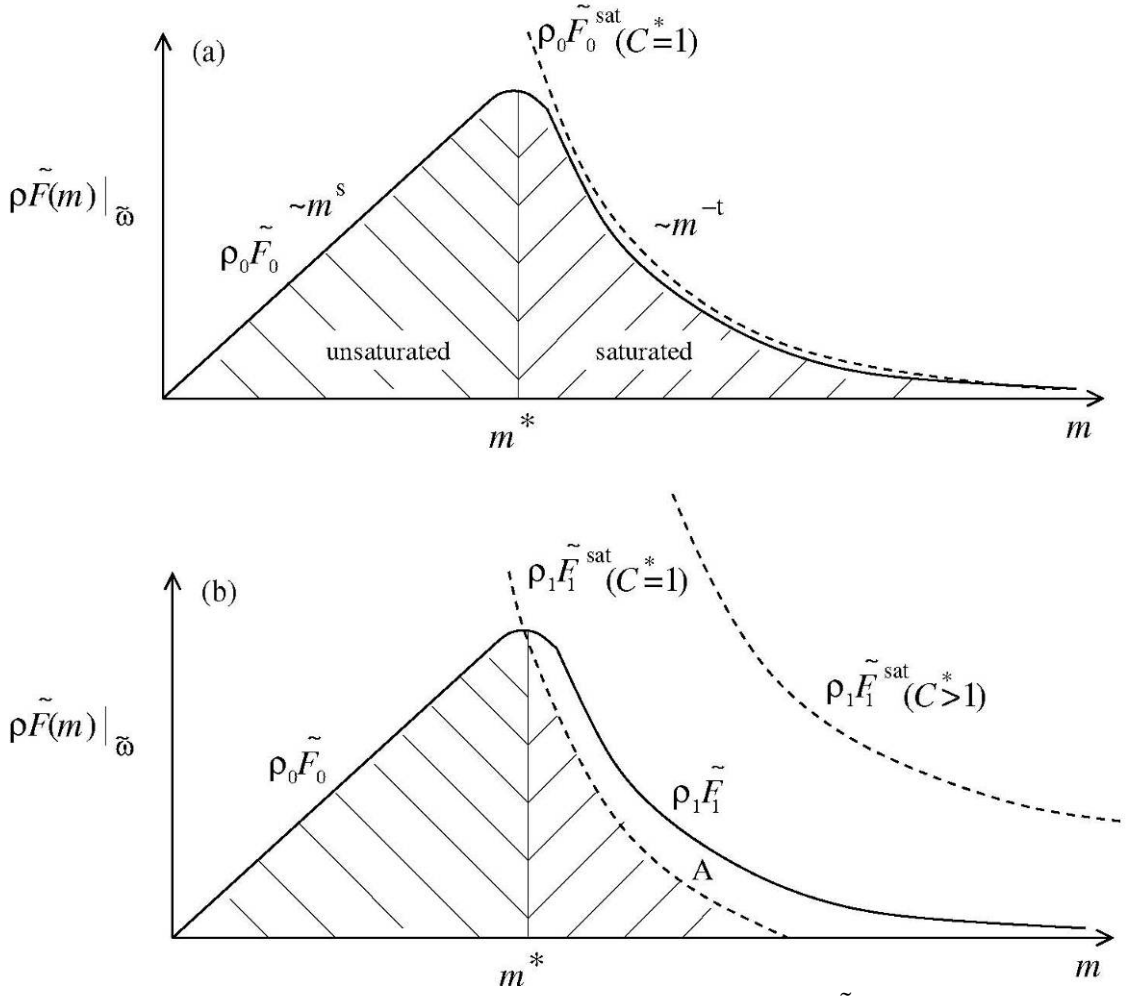

Figure 1: Schematic of the gravity wave momentum flux spectral density, $\rho\tilde{F}$, as a function of vertical wavenumber at constant intrinsic frequency $\tilde{\omega}$ at (a) $z = z_0$ and (b) $z_1 > z_0$. In (a) the launch spectrum, $\rho_0\tilde{F}_0$, consists of an unsaturated part proportional to m^s (for $m < m_*$) and a saturated part proportional to m^{-t} (for $m > m_*$). $\rho_0\tilde{F}_0^{sat}(C^*=1)$ equals $\rho_0\tilde{F}_0$ in the saturated part of the spectrum. In (b) the saturation curve $\rho_1\tilde{F}_1^{sat}(C^*=1)$ at height $z = z_1$ has intersected $\rho_1\tilde{F}_1$. $\rho_1\tilde{F}_1$ is set equal to $\rho_1\tilde{F}_1^{sat}(C^*=1)$, and the resulting momentum deposition is indicated by region A. For the case of $\rho_1\tilde{F}_1^{sat}(C^*>1)$, the saturated curve is offset vertically and does not intersect $\rho_1\tilde{F}_1$, indicating that nonlinear dissipation has not occurred. Therefore, increasing C^* increases the height at which momentum is deposited.

The momentum flux density launch spectrum is most easily illustrated for the case of $U(z)=0$ and constant N, in which case $\widetilde{\omega}=\omega$, $\widetilde{c}=c$, and $m=N/c$ are all independent variables, and $\rho_0\widetilde{F}_0(m,\widetilde{\omega},\phi)$ is conserved for conservative propagation. Figure 1(a) shows a schematic of the dependence on m of $\rho_0\widetilde{F}_0$ for a single azimuth and fixed $\widetilde{\omega}$, illustrating how the gravity wave spectrum is represented by an unsaturated part at small-*m* increasing with m^s, and a saturated part at large-*m* decreasing with m^{-t}, with m_* separating the two parts.

2.3 Critical level filtering and nonlinear dissipation

The two dissipative mechanisms which require modelling are critical level filtering and nonlinear dissipation. The application of critical level filtering follows from $\widehat{U}_0=0$ at launch level, which sets an absolute lower bound for critical level filtering of $\hat{c}=0$. If on the next level, $z_1(>z_0)$, U increases such that $\widehat{U}_1>\widehat{U}_0$, then waves with phase speeds in the range $\widehat{U}_0\le\hat{c}\le\widehat{U}_1$ encounter critical level filtering between these two levels. The momentum flux corresponding to these phase speeds is removed from $\rho\overline{F}(\hat{c},\phi)$ and deposited to the flow in this layer. The same procedure is repeated for subsequent layers and in all azimuths. The portions of the wave spectrum which survive critical level filtering at some level are allowed to conservatively propagate upwards to the next level, where they are checked for nonlinear dissipation.

The momentum deposited into the model layer during critical level filtering drags the mean flow towards the dissipating waves phase speed. However, depending on vertical resolution, there may not be enough mass to accept the momentum without forcing the mean flow past the phase speed of the waves. This is corrected by restricting the amount of momentum to be deposited so that the wind speed never exceeds the wave phase speed, with any excess momentum deposited in the layer below to conserve momentum.

The constraint, namely, that gravity waves cannot readily pass through critical layers, results in the filtering out of gravity waves orientated in the same direction as the background wind. Thus the westerly/easterly winter/summer stratospheric winds lead to a greater easterly/westerly bias in the gravity wave spectrum as the waves propagate through the stratosphere, and to larger net gravity wave easterly/westerly momentum fluxes entering the mesosphere (Garcia and Solomon, 1994). Therefore, the strong westerly zonal circulation of the southern winter leads to large net easterly momentum fluxes and strong easterly forcing at mesospheric altitudes. It follows that if the zonal winds in the stratosphere are weak, then neither westerly nor easterly propagating waves will be preferentially dissipated there. In this case the waves carry little net momentum flux into the mesosphere, so that when they are dissipated the net forcing will be small. This is often the case during spring and autumn as these seasons are associated with weak mean westerlies in both hemispheres (Andrews et al., 1987; Holton, 2004). This is also the reason why strong planetary wave driving in the northern winter stratosphere reduces the importance of gravity wave forcing, as it drives the stratosphere away from radiative equilibrium, reducing the strength of the westerly jet, and consequently the net easterly momentum flux into the mesosphere (Garcia and Solomon, 1994). Moreover, the strong dependence of mesospheric drag on filtering by the seasonally varying tropospheric and stratospheric winds justifies the use of a constant (and unrealistic) gravity wave source spectrum, in that the winds filter out any unrealistic gravity waves which are launched. This, however, requires that the gravity wave launch height is

suitably low, i.e. the upper troposphere (Manzini and McFarlane, 1998). Akmaev (2001) showed that overly strong winter westerlies can result in excessive easterly gravity wave drag in the mesosphere and unrealistically high polar stratopause temperatures.

The application of nonlinear dissipation is empirically defined by limiting the wave energy density at large-*m* to the observed m^{-3} dependence. This is achieved by specifying a saturation upper bound on the value of the wave energy density at each level, which in $m-\tilde{\omega}$ space is expressed as

$$\tilde{E}^{sat}(m,\tilde{\omega},\phi)=C^{*}B\left(\frac{m}{m_{*}}\right)^{-3}N^{2}\tilde{\omega}^{-p} \tag{18}$$

where C^{*} is a constant. Using (7), (8), and (9), and following the same procedure used to determine (16), this can be expressed in $\hat{c}-\varphi$ space in the form

$$\rho\overline{F}^{sat}(\hat{c},\phi)=\rho C^{*}A\frac{\hat{c}-\hat{U}}{N}\left(\frac{\hat{c}-\hat{U}}{\hat{c}}\right)^{2-p} \tag{19}$$

The constant C^{*} allows the normalization of the saturated spectrum to be specified independently of the launch spectrum (McLandress and Scinocca, 2005). Setting $C^{*}=1$ assumes that $\rho_0\overline{F}_0^{sat}$ asymptotically equals $\rho_0\overline{F}_0$ at large *m*, indicating that the tail of the launch spectrum is saturated [as shown by Figure 1(a)].

As discussed in section 1, to conserve the momentum flux density of spectral elements undergoing conservative propagation, the amplitude of the wave field must increase in response to diminishing density as the wave spectrum propagates vertically. Eventually the amplitude must become sufficiently large so that the wave field becomes unstable and nonlinear dissipation occurs. However, since the expression for $\rho\overline{F}(\hat{c},\phi)$ is conserved, it remains constant and unaffected by the decrease in density. The density effect is incorporated into the saturated momentum flux density, $\rho\overline{F}^{sat}(\hat{c},\phi)$, which is not conserved and reduces with height with the decreasing density. Therefore, all spectral elements that undergo conservative propagation are constrained by the saturation condition

$$\rho\overline{F}(\hat{c},\phi)\le\rho\overline{F}^{sat}(\hat{c},\phi) \tag{20}$$

Application of (20) implies for $\rho\overline{F}(\hat{c},\phi)>\rho\overline{F}^{sat}(\hat{c},\phi)$ that the momentum flux corresponding to these phase speeds is removed from $\rho\overline{F}(\hat{c},\phi)$ and deposited to the flow in this layer, and that $\rho\overline{F}(\hat{c},\phi)$ is set equal to $\rho\overline{F}^{sat}(\hat{c},\phi)$. The same procedure is repeated for subsequent layers and in all azimuths. Increasing $C^{*}(>1)$ implies that application of (20) will occur at a higher altitude where the density is smaller, increasing the height at which momentum is deposited (McLandress and Scinocca, 2005). In general, as the spectrum propagates to higher and higher elevations, the saturation condition is imposed on spectral elements at progressively larger values of $\hat{c}$.

The application of the saturation mechanism is again most easily illustrated for the case of $U(z) = 0$ and constant N. Figure 1(b) illustrates that at some height $z_1 (> z_0)$, the momentum flux density $\rho_1 \widetilde{F}_1$ remains identical to the launch value of $\rho_0 \widetilde{F}_0$ for a single azimuth and fixed $\widetilde{\omega}$. However, the decrease in the saturated momentum flux density with height as a consequence of decreasing density is represented by the downward shift of the $\rho_1 \widetilde{F}_1^{sat}$ curve, with the $C^* = 1$ curve intersecting the launch spectrum, resulting in an amount of momentum flux indicated by the area labelled 'A' being deposited to the flow in this azimuth. Also shown is the saturated curve $\rho_1 \widetilde{F}_1^{sat}$ for $C^* > 1$, which is offset vertically from $\rho_1 \widetilde{F}_1^{sat}$ for $C^* = 1$, and does not intersect the launch spectrum. Consequently, no momentum flux is deposited into the flow for this case. These two curves will eventually intersect at some higher elevation where the density is smaller.

After application of the dissipative mechanisms, the resulting momentum flux profiles are used to derive the net eastward, $\rho \overline{F}_E$, and northward, $\rho \overline{F}_N$, momentum flux. Defining U and V as the zonal and meridional wind components respectively, the resulting forcing terms or wind tendencies in each of these directions (i.e. the force per unit mass) is given by the vertical divergence of $\rho \overline{F}_E$ and $\rho \overline{F}_N$

$$F_u, F_v = \frac{\partial U, V}{\partial t} = g \frac{\partial \rho \overline{F}_E, \rho \overline{F}_N}{\partial p} \tag{21}$$

where F_u, F_v are the forcing terms in the zonal and meridional directions, g is the gravitational constant, t is time, and p is pressure.

2.4 Momentum conservation

At the model top, momentum conservation is achieved by depositing any remaining gravity wave momentum flux, such that here $\rho \overline{F}(\hat{c}, \phi) = 0$ (remembering that the net input momentum flux at launch level is zero). Shepherd and Shaw (2004) and Shaw and Shepherd (2007) suggest that if momentum is not conserved, then there is a non-negligible spurious downward influence in response to gravity wave drag feedbacks to planetary wave or radiative perturbations. Moreover, they suggest that by enforcing momentum conservation, the principal of 'downward control' implies that it is not necessary to explicitly simulate the region in which gravity wave drag occurs in order to represent the effects of that drag on regions below, reducing the sensitivity to model top height. Downward control implies that the gravity wave drag induced downwelling and dynamical heating at any altitude *z* depends on the gravity wave drag above that level, and that if momentum is conserved, then the gravity wave drag above *z* is determined by the momentum flux at *z* (Haynes et al., 1991). Relative to a simulation with a model top of 96 km, Lawrence (1997) had shown that the lack of the lower thermosphere in simulations with a 80 km model top resulted in a 40% underestimation of downwelling over the winter pole, and concluded that the model top height should be above some such height at which there is no longer any significant momentum flux being carried by the gravity waves [i.e. the lower thermosphere (Norton and Thuburn, 1999)]. However, Shaw and Shepherd (2007) suggest this strong sensitivity was because of non-conservation of momentum. This is convenient, as the IFS does not include the lower thermosphere as its model top is approximately 80 km at the mesopause (which separates the mesosphere from the thermosphere), and extending it to include this region is non-trivial. Firstly, the lower

thermosphere is very high above levels where the analysis is constrained by satellite observations (McNally, 2003). Secondly, the current IFS radiation scheme (Morcrette et al., 2007) does not represent non-local thermodynamical equilibrium, which is an important contributor to upper atmosphere long-wave radiative forcing. It is further suggested by Shepherd and Shaw (2004) that enforced momentum conservation increases the robustness of the parametrized gravity wave drag to changes in the source spectrum, background flow, and the chosen method of gravity wave dissipation.

2.5 Coordinate stretch

McLandress (1998) discusses that it is the small-*m* portion of the launch spectrum, which is associated with relatively large values of momentum flux, which is most crucial to the gravity wave drag in the mesosphere. By contrast, the large-*m* portion is associated with relatively small values of momentum flux, and a large proportion of these waves are removed by the dissipation mechanisms soon after the gravity wave spectrum is launched. Thus, higher resolution at large phase speeds (i.e. small-*m*) would improve the accuracy of the calculation of gravity wave drag, as well as reduce the computational cost. In the SO3 scheme this was achieved by applying a coordinate stretch to the discretized values of $\hat{c}$ in each of the n_ϕ azimuths. The untransformed variable, X, is given as

$$X = \frac{1}{\hat{c}} \tag{22}$$

which from (3) is proportional to the launch vertical wavenumber. The transformed variable, $\breve{X}$, has uniform resolution, $d\breve{X}$, which is defined by

$$X = B_1 e^{(\breve{X} - X_{\min})/\Gamma} + B_2 \tag{23}$$

and

$$\frac{dX}{d\breve{X}} = \frac{B_1}{\Gamma} e^{(\breve{X} - X_{\min})/\Gamma} \tag{24}$$

where $X_{\min} \le X \le X_{\max}$, $X_{\min} = 1/\hat{c}_{\max}$, $X_{\max} = 1/\hat{c}_{\min}$, $dX = (X_{\max} - X_{\min})/n_{\hat{c}}$, and Γ is the half-width of the stretch, and $\hat{c}_{\min}$ and $\hat{c}_{\max}$ are respectively the minimum and maximum launch phase speed in each azimuth. Imposing the constraint $X_{\min} \le \breve{X} \le X_{\max}$ implies

$$B_1 = \frac{X_{\max} - X_{\min}}{e^{(X_{\max} - X_{\min})/\Gamma} - 1} \tag{25}$$

and

$$B_2 = X_{\min} - B_1 \tag{26}$$

This transformation is independent of azimuth and geographic location, and so need only be computed once.

3. Parameter settings and design of experiment

3.1 Parameter settings

The SO3 scheme requires specification of the following free parameters:

- p (= 1 or 1.5): the $\widetilde{\omega}$ exponent in (6) and (18).
- s (= -1 or 0 or 1): small-m spectral slope in (6) [see Figure 1].
- $\rho_0 \overline{F}_0^{total}$: total launch momentum flux density in each azimuth (Pa).
- p_{launch} : launch elevation (hPa).
- n_ϕ (= 4 or 8 or 16 ...): number of discrete equally spaced azimuths.
- B: non-dimensional constant in (6).
- C^* : non-dimensional constant in (18).
- m_* : the transitional vertical wavenumber in (6) and (18) (units of m^{-1}).

The values of p are restricted to integer and half-integer values (i.e. 1 or 1.5) to reduce the computational cost (Scinocca, 2002). The values of s are restricted to -1 or 0 or 1.

Ern et al. (2006) compared absolute values of gravity wave momentum flux derived from CRISTA-1 and CRISTA-2 satellite measurements with values simulated using the 'ultra-simple' hydrostatic non-rotating parametrization of Warner and McIntyre (2001). The findings suggested that the highest correlation with the observed momentum flux was with p_{launch} between 700 and 450 hPa, with the 450 hPa launch level giving the best results. Comparison with the standard choice of launch parameters used by the scheme (e.g. $s = 1$ and $m_* = 2\pi / 2000\,\mathrm{m}^{-1}$) showed good agreement with the observed momentum flux. However, the optimum value of m_* was suggested to be $2\pi / 2500\,\mathrm{m}^{-1}$ for $s = 1$. Scinocca (2003) demonstrated a close correspondence in zonal-mean zonal winds for January and July between the non-hydrostatic rotational Warner and McIntyre (1996) scheme and the SO3 scheme using parameter settings of $p = 1.5$, $s = 0$, $m_* = 2\pi / 2000\,\mathrm{m}^{-1}$, p_{launch} = 125 hPa, $\rho_0 \overline{F}_0^{total} = 4.25 \times 10^{-4}$ Pa, and $n_\phi = 4$ (with cardinal directions of N, S, E, W).

3.2 Design of experiment

Based on this information, and some initial sensitivity/optimization experiments, the set of parameters chosen for the SO3 simulation were $p = 1$, $s = 1$, $m_* = 2\pi / 2000\,\mathrm{m}^{-1}$, p_{launch} = 450 hPa, $\rho_0 \overline{F}_0^{total} = 4.0 \times 10^{-3}$ Pa, and $n_\phi = 4$ (with cardinal directions of N, S, E, W).

Preliminary tests of IFS cycles 32R3, 33R1, and 33R1 (+ CLIM_GHG), showed that cycle 33R1 (+ CLIM_GHG) gave the best representation of winter temperatures and the winter westerly jet. CLIM_GHG is an improved greenhouse gas (GHG) climatology of O_3, CO_2, CH_4, N_2O, CFC11, CFC12, CFC22, and

CC14 from the MOBIDIC 2D chemical transport model (Cariolle et al., 2008). Therefore, cycle 33R1 (+ CLIM_GHG) was chosen as the basis for all experiments, since improvements to the circulation are important for the implementation of the SO3 scheme as it relies on realistic winds to filter the upward propagating (unrealistic) gravity wave source. Cycle 33R1 (+ CLIM_GHG) is the RF simulation. The SO3 simulation is 33R1 (+ CLIM_GHG), but with the SO3 scheme in place of Rayleigh friction.[1]

The resolution of the experiments is T159L91, which is characterised by a horizontal resolution of around 125 km and a model top located at 0.01 hPa (i.e. the mesopause). The 91 vertical levels are arranged with approximately 53 in the troposphere (i.e. between the surface and 100 hPa), 29 in the stratosphere (i.e. between 100 and 1 hPa), and 9 in the mesosphere (i.e. between 1 and 0.01 hPa). Taking the respective heights of the tropical tropopause, stratopause, and mesopause, as 20, 50 and 80 km, gives an approximate vertical resolution of ~1 km in the tropical stratosphere and ~3 km in the tropical mesosphere. Due to the large inter-annual variability of the atmosphere, and in particular the northern hemisphere atmosphere (Holton, 2004), both the RF and SO3 simulations comprise eight 13 month integrations covering years 1994 to 2001 and beginning on November 1 of the preceding year (i.e. 1993 to 2000), which are then used to evaluate the (8-year) mean model climate. The integrations are forced at the lower boundary by observed sea-surface temperatures.

From section 2.5, the SO3 scheme also requires specification of $n_{\hat{c}}$, $\hat{c}_{\max}$, $\hat{c}_{\min}$, and Γ. To determine the sensitivity to $n_{\hat{c}}$ and $\hat{c}_{\max}$, T159L91 simulations comprising eight integrations covering years 1994 to 2001, but beginning in each case on July 1 and lasting one month, were conducted. Here, $\hat{c}_{\min}$ (= 0.25 m s^{-1}) and Γ [= 0.25 (m s)$^{-1}$] were constant. The combinations of $(\hat{c}_{\max}, n_{\hat{c}})$ examined were (100 m s^{-1}, 100), (200 m s^{-1}, 100), (100 m s^{-1}, 50), and (200 m s^{-1}, 50). Vertical profiles of the resulting average T + 720 h mean winter-time (i.e. averaged between 60 and 90°S) zonal gravity wave drag are shown in Figure 2. It is clear that a choice of $\hat{c}_{\max}$ = 100 m s^{-1} and $n_{\hat{c}}$ = 50 is sufficient, with the lower resolution beneficial for reducing the cost of the scheme. Nevertheless, as computational cost was not an issue in this study, the values chosen for the SO3 experiment detailed here are $\hat{c}_{\max} = 100$ m s^{-1} and $n_{\hat{c}} = 100$, with additional standard values of $\hat{c}_{\min} = 0.25$ m s^{-1} and $\Gamma = 0.25$ (m s)$^{-1}$. It should be noted that Scinocca (2003) showed that the SO3 scheme can be run accurately at a resolution as low as $n_{\hat{c}} = 15$.

[1] After the preparation of this report it was discovered that there was a minor bug in the specification of the greenhouse gas concentrations of the CLIM_GHG climatology (correction of this bug reduced the impact of GHG). The conclusions in this report about the impact on the IFS middle atmosphere climate when gravity wave forcing is represented by the SO3 scheme are believed to be still valid, because both the control (RF) and non-orographic gravity wave experiment (SO3) use the same (but slightly incorrect) greenhouse gas concentrations.

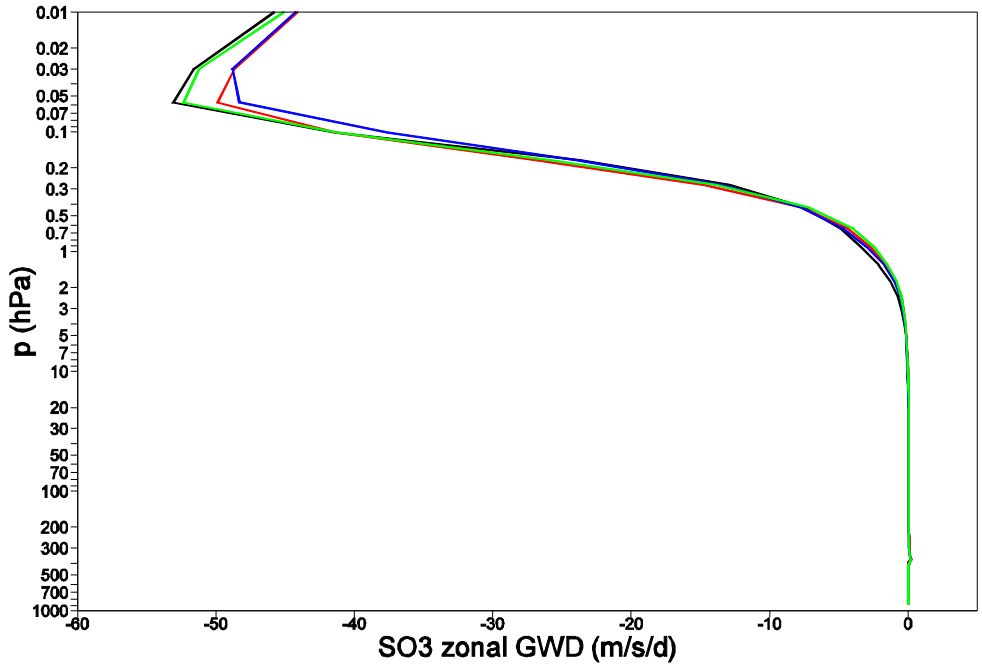

Figure 2: *Profiles of mean winter-time July 1 T + 720 h zonal-mean zonal gravity wave drag (m s^{-1} d^{-1}) averaged between 60 and 90°S for combinations of* $(\hat{c}_{\max}, n_{\hat{c}})$ *of (100 m s^{-1}, 100; black line), (200 m s^{-1}, 100; red line), (100 m s^{-1}, 50; blue line), and (200 m s^{-1}, 50; green line). The simulation results are 8-year averages. The vertical axis is pressure (hPa) and extends to 0.01 hPa.*

The RF and SO3 simulations parametrize momentum deposition from orographic gravity waves (Lott and Miller, 1997). Cycle 33R1 linearly smooths/spreads the orographic gravity wave momentum flux at an altitude of 0.1 hPa between 0.1 hPa and the model top. This was designed to reduce large and spurious orographic gravity wave drag (OGWD) values which occurred at the uppermost model levels (partly as a response of the requirement for conservation of momentum which implies the deposition of any remaining gravity wave momentum flux at the model top), which significantly degraded the model climate. However, Orr and Wedi (2009) re-assessed the smoothing, and showed that it is beneficial to the model climate if it is switched off, as has been done here.

Rayleigh friction acts on both the zonal mean flow (i.e. zero zonal wave number) and waves (finite zonal wave number). In the SO3 simulations, the gravity wave scheme replaces Rayleigh friction on the mean flow, which requires Rayleigh friction to be turned off. Another reason for requiring Rayleigh friction on the mean flow to be turned off is that Shepherd et al. (1996) showed that its long-term response is to couple unrealistically with any imposed local force, such as additional parametrized gravity wave drag. However, initial tests turning off Rayleigh friction on the mean flow produced an unrealistic spurious jet in the tropical stratosphere due to the relatively coarse vertical resolution at this level preventing small scale resolved waves from propagating vertically. These waves would break at this level and deposit momentum, causing the jet. To minimize this feature, Rayleigh friction on the zonal mean flow therefore remained on in the SO3 experiment, but the formulation of (1) was modified to be heavily damped with respect to latitude and constrained to a narrow region over the tropics, such that

$$F_u = -K(z)U\cos^q(\varphi) \qquad (27)$$

where φ is the latitude and $q = 12$. Rayleigh friction acting on waves remained on. Both simulations include enhanced horizontal diffusion towards the top of the model, which along with Rayleigh friction on the waves, acts to damp upward propagating resolved waves which are (unrealistically) reflected from the model top, limiting their penetration into the model domain.

The remainder of this technical memorandum predominately discusses monthly mean results for January and July, computed from fields at 24-h intervals. Comparing these fields with monthly mean results based on 6-h intervals showed negligible differences. The simulated results are compared to observations, comprising SPARC climatology (Randel et al., 2004) and ECMWF ERA-Interim (ERAI) reanalysis. The SPARC climatology provides a dataset up to a height of 0.01 hPa, from sources such as reanalysis, satellite measurements, rocketsonde wind and temperature data, and lidar temperature measurements. However, as there are relatively few measurements directly over the polar regions, the climatology effectively stops at 80^{o} north and south, with values at higher latitudes interpolated. ERAI uses a 12 hour 4D-VAR, T255L60 assimilating model (i.e. a horizontal resolution of around 60 km and a model top of 0.1 hPa) based on cycle 31R1 of the IFS. Equivalent January and July mean 8-year ERAI fields are computed for the 1994 to 2001 period of the simulations. As is shown in the next section, ERAI data does not capture the expected warm (separated) polar stratopause in either the northern or southern winter, suggesting that the wind and temperature data at these levels should be treated with caution, and which perhaps reflects the paucity of satellite data used in the reanalysis at these altitudes and high latitudes, and/or biases in the assimilating model.

4. Changes in the temperature structure and circulation

4.1 January

The average zonal-mean temperature and zonal wind fields obtained from the RF and SO3 simulations are shown in Figure 3 for January. The SO3 and RF temperature fields differ significantly in the northern winter polar stratosphere and stratopause, and the southern summer polar upper mesosphere. The SO3 simulation shows a warmer/colder northern winter polar stratosphere/stratopause, and a colder southern summer polar upper mesosphere. The SO3 and RF zonal wind fields also differ, with the SO3 simulation showing stronger southern summer easterlies and northern winter westerlies. The SPARC and ERAI zonal-mean temperature and zonal wind fields are given for comparison. The SO3 simulation of the southern summer hemisphere agrees well with the observations, despite the easterly jet not quite capturing the required poleward slant of its edge with height. The cold southern summer polar upper mesosphere is suggestive of strong upwelling at the summer pole being driven by a realistic overturning circulation through the inclusion of strong gravity wave drag. However, the warm (separated) northern winter polar stratopause apparent in the SPARC data is not reproduced by the SO3 simulation. This feature is also erroneously not captured by the ERAI data. Additionally, the northern winter polar stratosphere of the SO3 simulation is slightly warmer than observed. Despite these discrepancies, the agreement between the SO3 simulation and observations is more satisfactory than that obtained with the RF simulation. The RF simulation is characterised by a temperature structure which is unrealistically close to radiative equilibrium, i.e. an excessively warm southern summer polar upper mesosphere and an excessively cold northern winter polar stratosphere (relative to the SPARC data), which is consistent with the drag supplied by Rayleigh friction being insufficient and the development of an overly weak overturning circulation (Orr and Wedi, 2009).

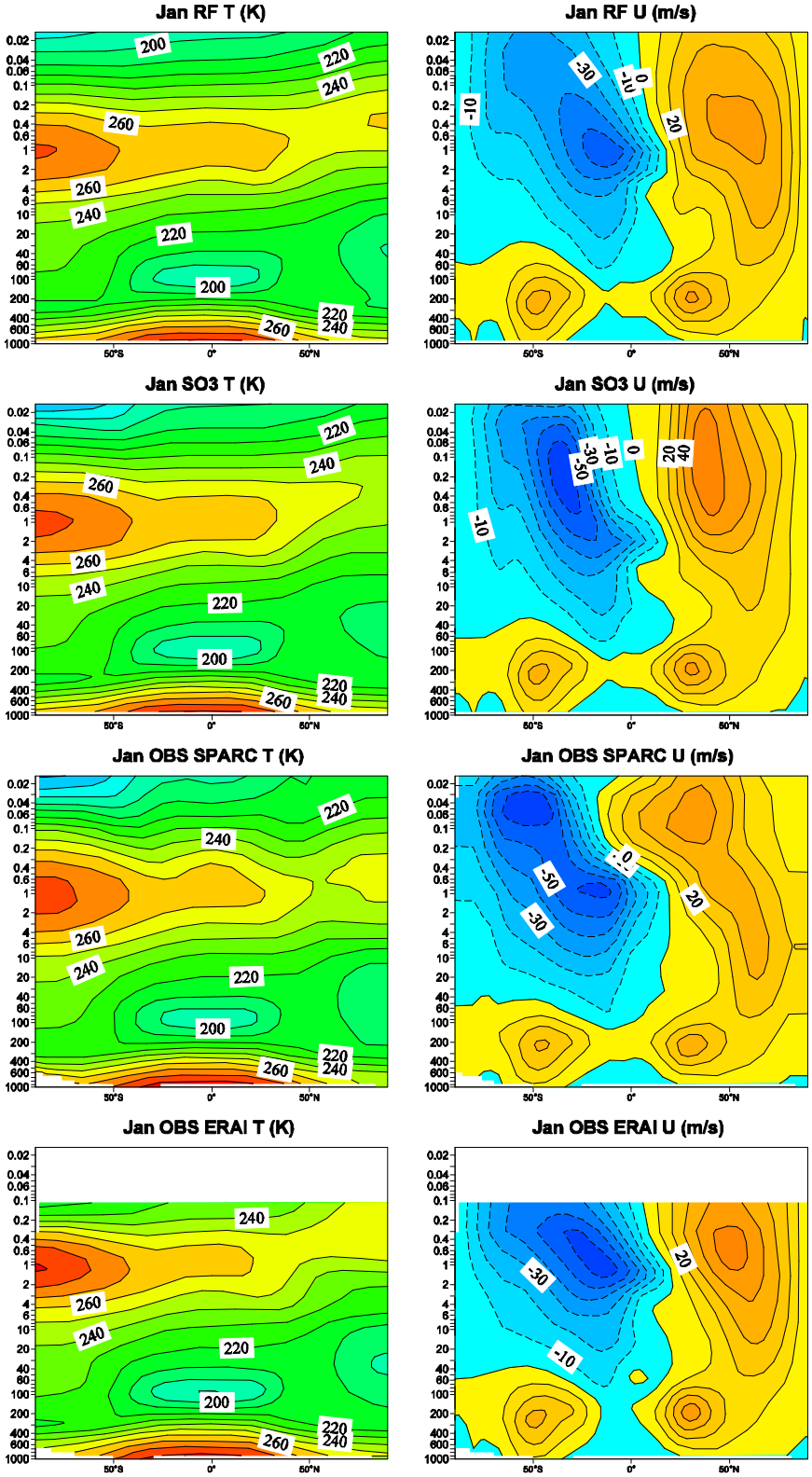

Figure 3: *Average January cross-sections of zonal-mean temperature (left) and zonal wind (right) for the RF (top) and SO3 (second) simulations, and SPARC (third) and ERAI (lower) observations. Simulation results are 8-year averages. SPARC data north and south of 80°N and 80°S respectively is constant and interpolated. ERAI data stops at 0.1 hPa. Contour intervals are 10 K and 10 m s^{-1}. The vertical axis is pressure (hPa) and extends to 0.015 hPa.*

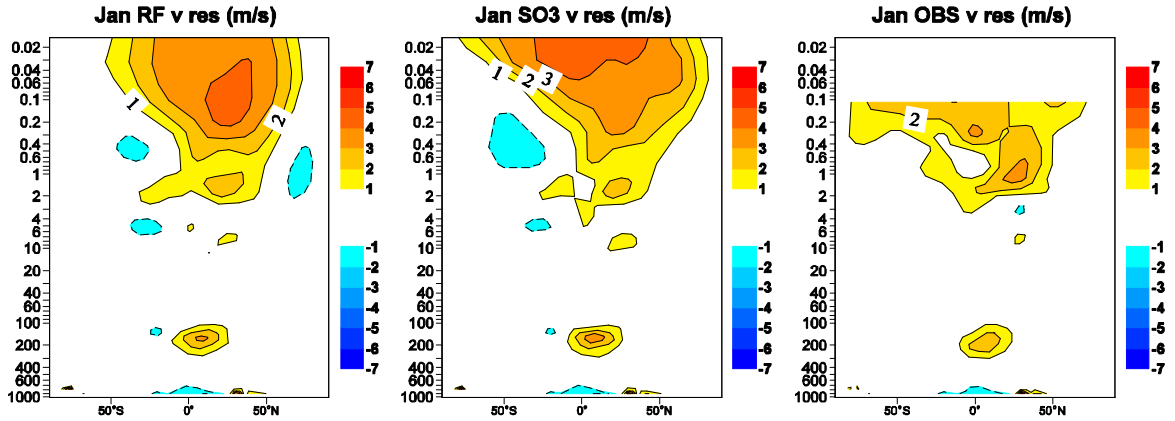

Figure 4: *Average January cross-sections of zonal-mean residual meridional velocity (m s^{-1}) for the RF (left) and SO3 (middle) simulations, and observations (right). The simulation results are 8-year averages. The observations are ERAI data up to 0.1 hPa. The contour interval is ± 1, 2, 3, 4, 5, 6, 7 m s^{-1}. The vertical axis is pressure (hPa) and extends to 0.015 hPa.*

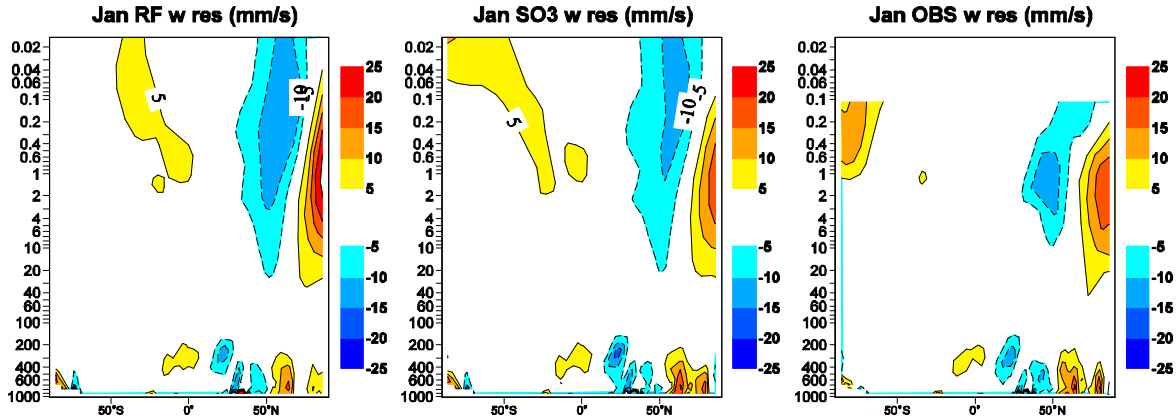

Figure 5: *As for Figure 4, but showing the residual vertical velocity. The contour interval is ± 5, 10, 15, 20, 25 mm s^{-1}.*

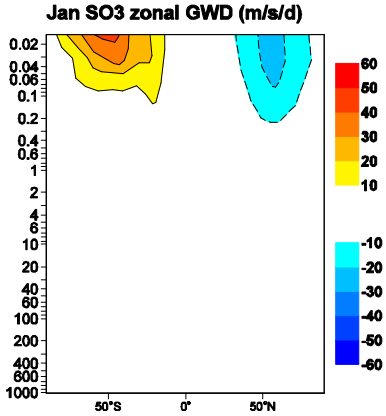

Figure 6: *Average January cross-sections of zonal-mean zonal gravity wave drag (m s^{-1} d^{-1}) for the SO3 simulation. The simulation results are 8-year averages. The contour interval is ± 10, 20, 30 … 60 m $s^{-1}d^{-1}$. The vertical axis is pressure (hPa) and extends to 0.015 hPa.*

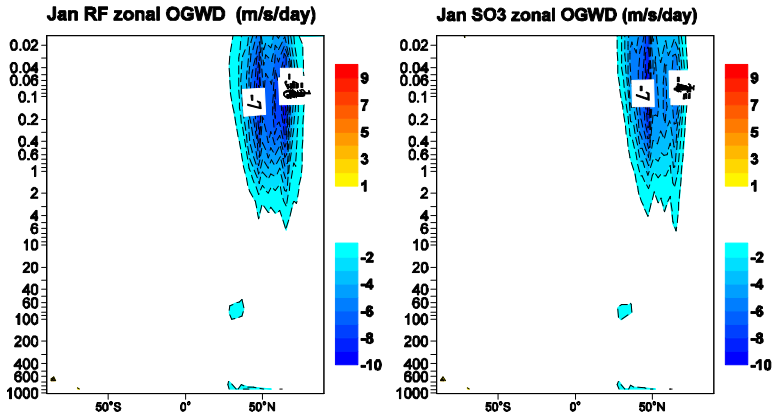

Figure 7: Average January cross-sections of zonal-mean zonal orographic gravity wave drag (m s^{-1} d^{-1}), from the RF (left) and SO3 (right) simulations. The simulation results are 8-year averages. The contour interval is ± 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 m s^{-1} d^{-1}. The vertical axis is pressure (hPa) and extends to 0.015 hPa.

Figure 4 shows the average zonal-mean residual meridional velocity obtained from the RF and SO3 simulations for January. Although both simulations show peak values reaching 4 m s^{-1}, the SO3 simulation shows flow extending as far as both poles and stronger flow in the upper mesosphere, which presumably reflects the influence of the stronger parametrized gravity wave drag. For comparison, ERAI data is also shown. This seems in better agreement with the SO3 field (i.e. flow is shown extending to both poles), however, as Figure 3 suggests, there is a lack of confidence in the ERAI data at these high altitudes and/or high latitudes.

Figure 5 shows the average zonal-mean residual vertical velocity obtained from the RF and SO3 simulations for January. Again, ERAI data is shown for comparison. The SO3 simulation reproduces the magnitude of the observed northern winter mid-latitude downwelling and polar upwelling, which are overestimated by the RF simulation (and in particular the polar upwelling). The SO3 simulation shows a stronger southern summer polar upwelling compared to that of the RF simulation, which again reflects the influence of strong gravity wave drag.

Figure 6 displays the SO3 average zonal-mean zonal gravity wave drag for January. No equivalent values are shown for the RF simulation. The peak northern winter value is -20 m s^{-1} d^{-1} in the upper mesosphere and is consistent with the findings of Smith and Lyjak (1985) who estimated gravity wave drag from the zonal momentum budget computed from seven months of satellite observations. Estimating the forcing by gravity waves as the residual, they calculated that values of about -20 m s^{-1} d^{-1} were required to obtain momentum balance during northern winter. Although this value was observed to occur in the lower mesosphere, suggesting that the SO3 simulation places the drag too high. However, the particularly coarse vertical resolution in the mesosphere (discussed in section 3) suggests that the height at which the parametrized drag is located is treated rather insensitively. The peak value in the southern summer mesosphere is 40 m s^{-1} d^{-1}.

The average zonal-mean zonal orographic gravity wave drag is shown in Figure 7 for January. Both the RF and SO3 simulations show areas of drag extending throughout the northern winter upper stratosphere and mesosphere, with peak values of around -8 m s^{-1} d^{-1}. The high altitude of the drag is perhaps surprising as

orographic gravity waves are typically limited to the troposphere and stratosphere (Eckermann and Preusse, 1999).

4.2 July

The average zonal-mean temperature and zonal wind fields for July are shown in Figure 8. The SO3 results show significant differences from those of the RF simulation, in particular: a warmer southern winter polar stratosphere, more horizontally aligned temperature contours immediately below the separated winter stratopause, a colder northern summer polar upper mesosphere, a weaker and closed westerly jet with a tropicward slant, located at a lower altitude, and a stronger easterly jet, all of which agree well with the observations and are consistent with a realistic overturning circulation being driven by the inclusion of strong gravity wave drag. The ERAI data again does not capture the warm separated southern winter polar stratopause. As with the equivalent January results, the RF extratropical circulation and temperature structure are unrealistically close to radiative equilibrium.

Figure 9 shows the average zonal-mean residual meridional velocity obtained from the RF and SO3 simulations for July. In contrast to the RF simulation, the SO3 values extend over the northern summer pole, which agrees with the ERAI data included for comparison. This is consistent with stronger SO3 residual vertical velocity over this pole (Figure 10), which similarly agrees with the observations. Over the southern summer pole, the SO3 simulation shows both stronger meridional and vertical velocities relative to the RF simulation, confirming that the improved SO3 circulation and temperature structure is due to the overturning circulation having strengthened. The peak SO3 residual meridional velocity is -6 m s^{-1}. Garcia and Solomon (1985) cite a large number of observational studies that show that the magnitude of the zonal-mean meridional velocity field is around 10 m s^{-1}.

Figure 11 displays the SO3 average zonal-mean zonal gravity wave drag for July. No equivalent values are shown for the RF simulation. Peak values of around -50 m s^{-1} d^{-1} occur in the southern winter mid to upper mesosphere, and of around 30 m s^{-1} d^{-1} in the northern summer upper mesosphere. Pulido and Thuburn (2006) used variational data assimilation principles to estimate gravity wave drag from Met Office global middle atmosphere analyses. For a one week period starting on July 2002, they estimated that the peak southern winter drag was around -50 m s^{-1} d^{-1}, in excellent agreement with the SO3 estimates. However, this peak value was observed to occur in the lower mesosphere, suggesting again that the SO3 simulation places the drag too high.

The average zonal-mean zonal orographic gravity wave drag for July is shown in Figure 12, indicating that the SO3 values are considerably smaller than the RF values. The improved closure of the SO3 simulated winter westerly jet (Figure 8) results in a (more realistic) reduction in momentum deposition at the model top due to earlier (lower down) orographic wave breaking as a consequence of the accompanying weaker zonal winds at the model top (Palmer et al., 1986). Comparing Figure 12 with Figure 7 shows that, as expected, southern winter orographic gravity wave drag values are smaller than those in the northern winter.

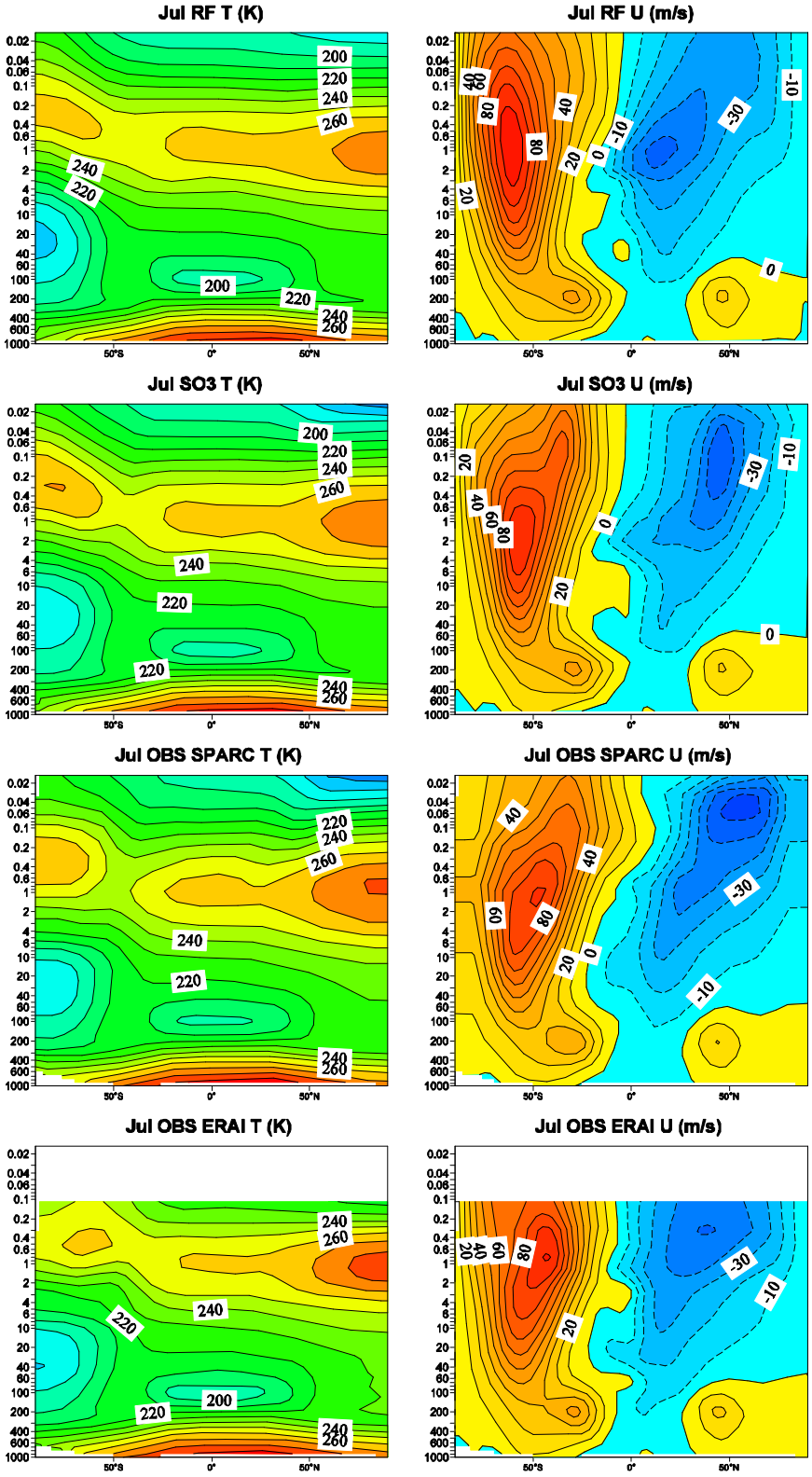

Figure 8: *As Figure 3, but for July.*

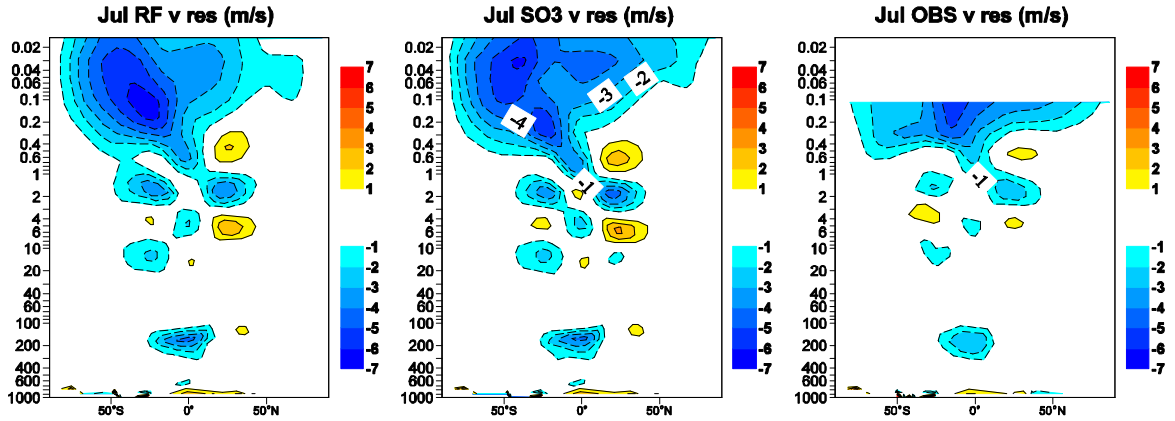

Figure 9: As Figure 4, but for July.

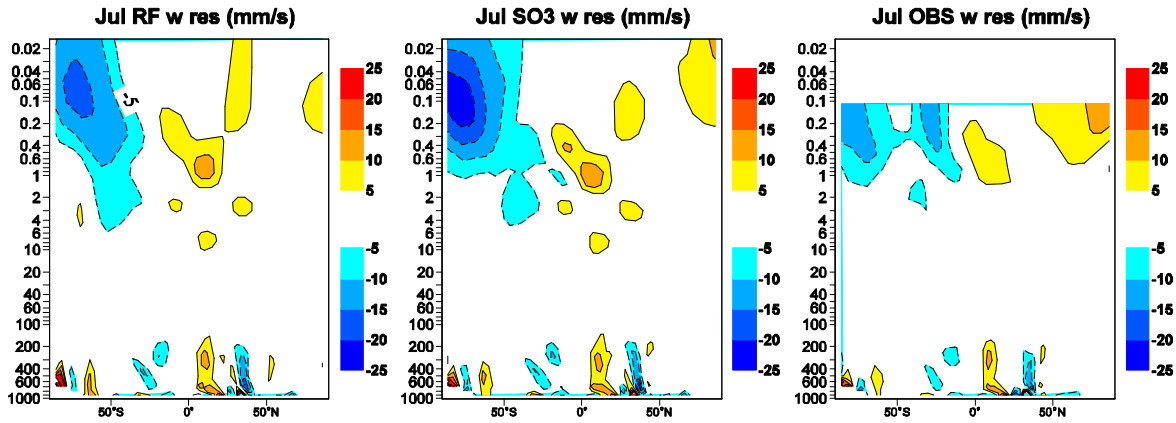

Figure 10: As Figure 5, but for July.

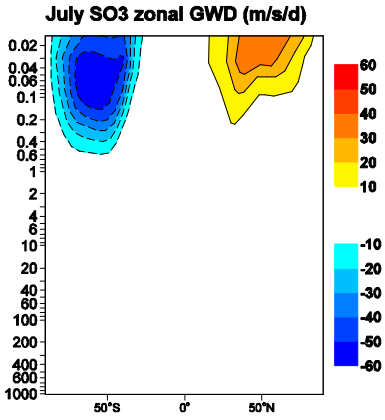

Figure 11: As Figure 6, but for July.

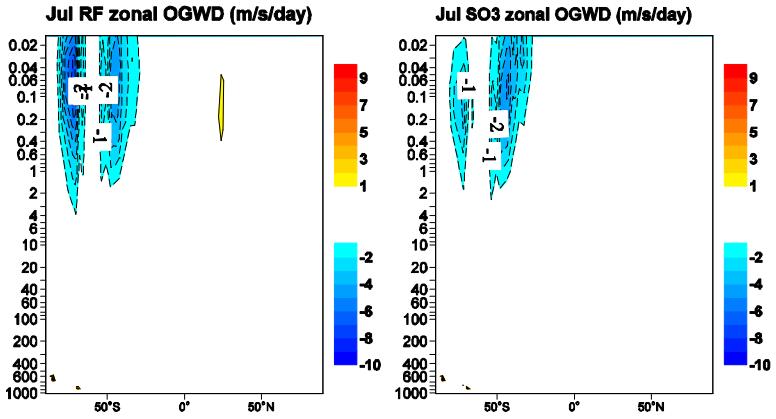

Figure 12: As Figure 7, but for July.

5. Changes in planetary waves

Boville (1986) demonstrated that the weaker northern winter jet is maintained by a positive feedback mechanism, such that: with weak background flow, planetary waves are able to penetrate more easily into the stratosphere, and are then deflected poleward rather than equatorward, resulting in stronger wave-mean flow interaction in the lower and middle stratosphere, maintaining the weak winds while dissipating the waves. By contrast, the stronger southern winter jet is consistent with weak planetary wave driving (e.g. Shine, 1989).

5.1 January

Figure 13 shows the stationary (northern hemisphere) planetary wave amplitude for January, in terms of zonal velocity anomaly at pressure levels of 100, 10, and 1 hPa, for the SO3 and RF simulations. For observations ERAI data is shown. The SO3 amplitude is in much better agreement with the observations than those of the RF simulation. Although at 1 hPa the SO3 planetary wave amplitude is slightly smaller than observed. The RF amplitude at 10 and 1 hPa is much is too small, suggesting excessive dissipation of the waves in the stratosphere.

Figure 14 shows the average January zonal-mean planetary wave forcing given by the stationary wave Eliassen-Palm (EP) flux divergence, confirming the good agreement between observed (ERAI data) and SO3 simulated amplitudes. It could be argued that relative to the RF simulation that the SO3 simulation shows an increase in easterly (negative) EP flux divergence just below the level of the stratopause (which based on the temperature structure of Figure 3 is around 0.3 hPa). Matsuno and Nakamura (1979) showed that associated with such an increase in wave driving would be increased sinking/rising of air below/above the dissipation region. The increase in rising air would cool the stratopause level in the SO3 simulation, limiting the warming induced by gravity wave drag produced downwelling, and perhaps explaining the relatively cold northern winter polar stratopause apparent in Figure 3. The excessive dissipation of the RF planetary waves in the stratosphere is reflected in reduced planetary wave driving in the mesosphere (i.e. excessively small EP-flux divergence values), as a consequence of there being less of the wave remaining to dissipate in this region.

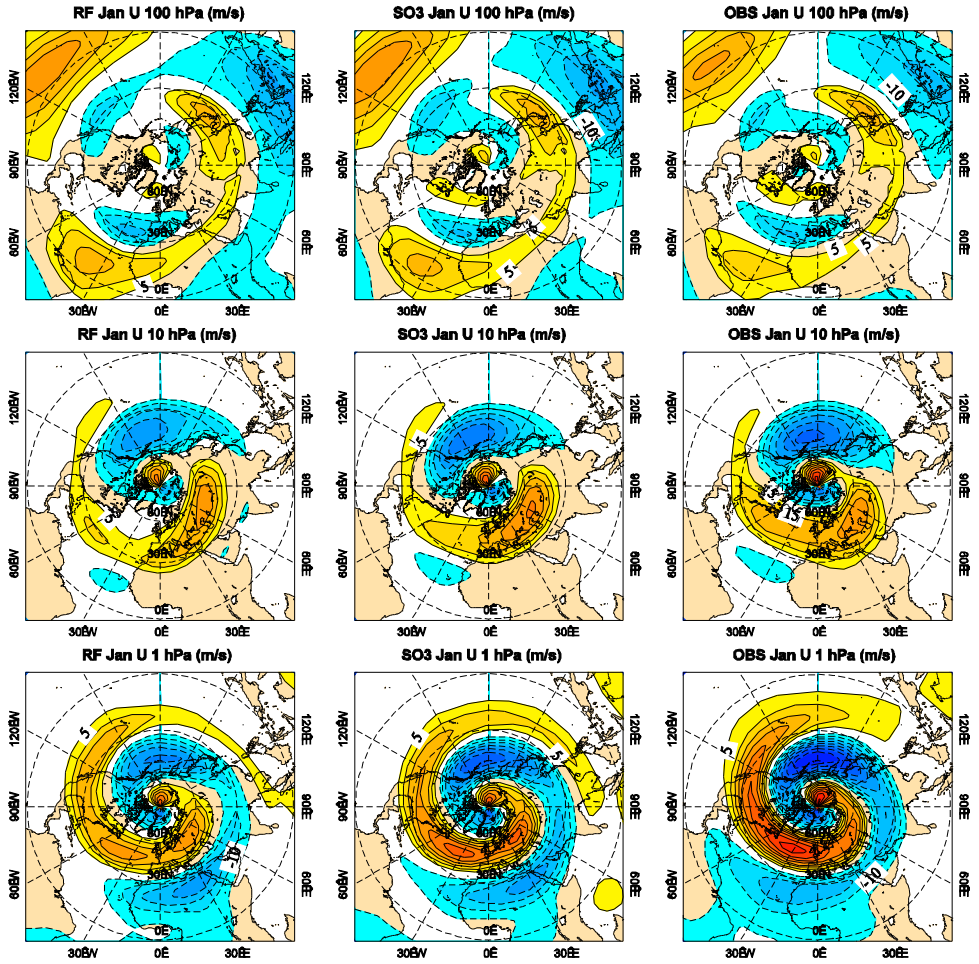

Figure 13: Average January (northern hemisphere) mean stationary planetary wave zonal wind amplitude (m s^{-1}) at 100 hPa (upper), 10 hPa (middle), and 1 hPa (lower), for the RF (left) and SO3 (middle) simulations, and observations (right). The simulation results are 8-year averages. The observations are ERAI data. The contour interval is ± 5, 10, 15 … 40 m s^{-1}.

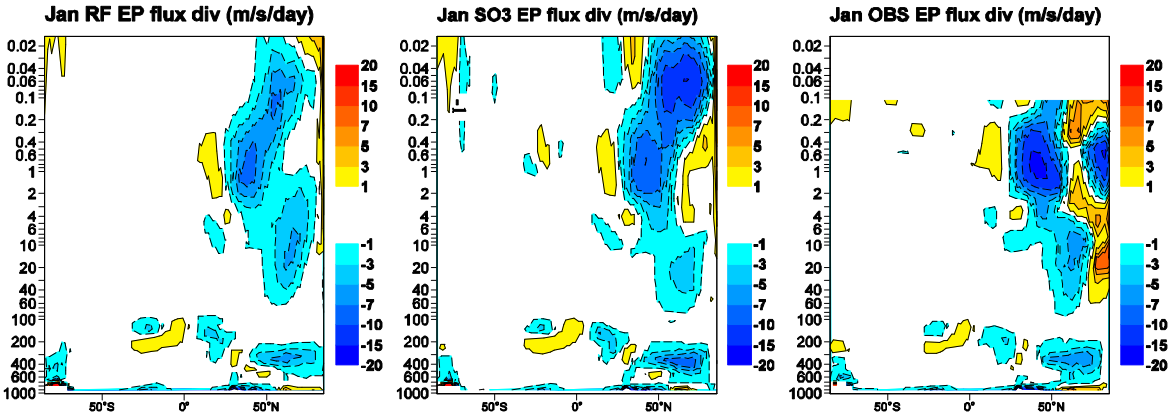

Figure 14: Average January cross-sections of zonal-mean stationary wave EP flux divergence (m s^{-1} d^{-1}) for the RF (left) and SO3 (middle) simulations, and observations (right). The EP flux is divided by acosφ, where a is the earth's radius and φ is the latitude. The simulation results are 8-year averages. The observations are ERAI data up to 0.1 hPa. The contour interval is ± 1, 2, 3, 5, 7, 10, 15, 20 m s^{-1} d^{-1}. The vertical axis is pressure (hPa) and extends to 0.015 hPa.

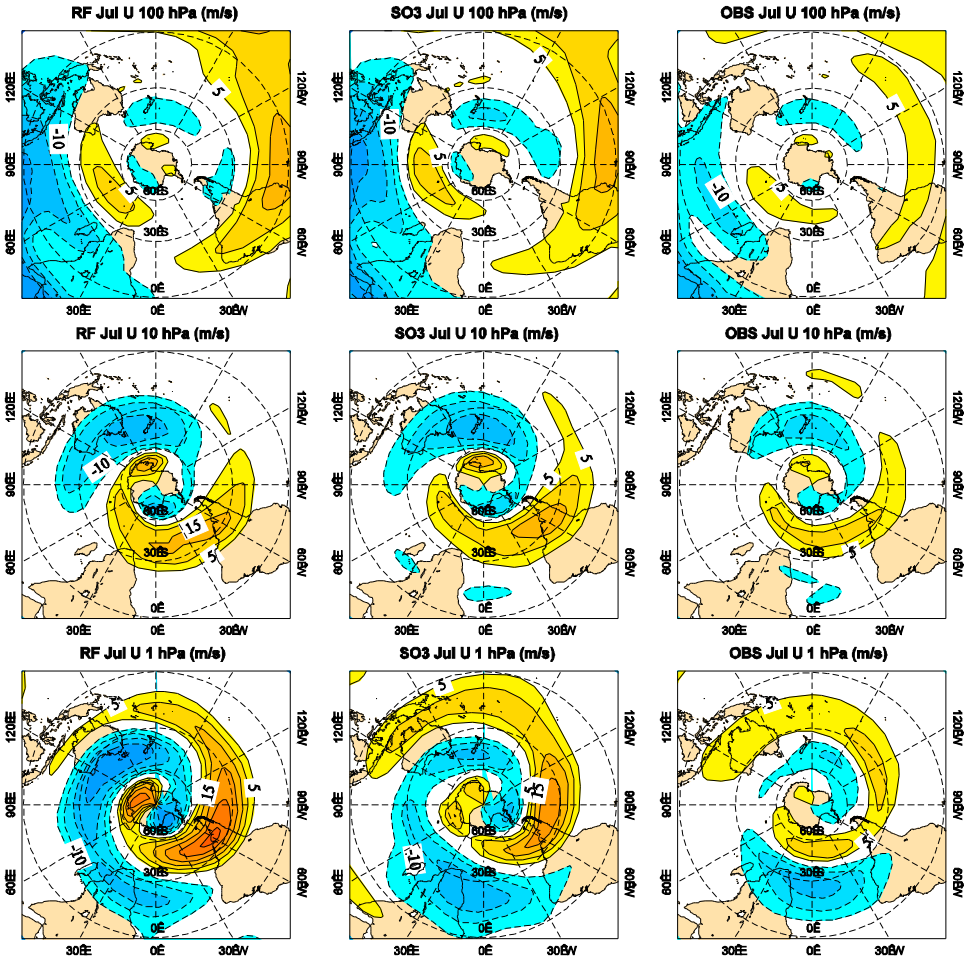

Figure 15: As Figure 13, but for July.

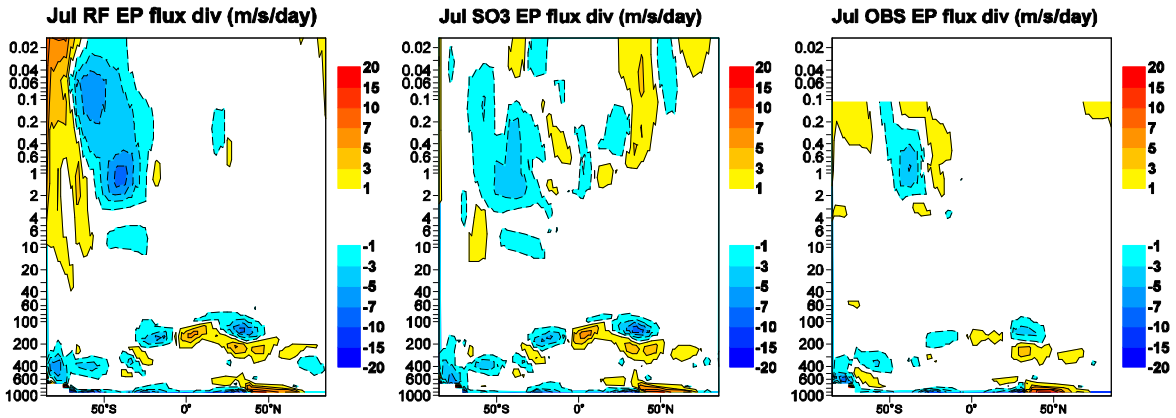

Figure 16: As Figure 14, but for July.

5.2 July

Figure 15 is analogous to Figure 13, but showing the stationary (southern hemisphere) planetary wave amplitude for July. The SO3 simulated amplitude at 1 hPa is markedly smaller than the RF amplitude, and in much better agreement with the observations. The increased dissipation of the SO3 wave verifies the conclusion of Manzini and McFarlane (1998) that planetary wave dissipation is enhanced by the introduction

of strong gravity wave drag (i.e. through the resulting weaker jet). Note that a feedback mechanism exists, such that if the enhanced dissipation of the SO3 simulation was to further weaken the jet, then the gravity wave drag would decrease due to reduced filtering of the gravity wave spectrum. By contrast, the RF amplitude is too large at 1 hPa, which is consistent with the argument of Boville (1986) that an overly strong background flow (Figure 8) is associated with insufficient dissipation.

The improvement in the SO3 simulation of the stationary planetary wave amplitude is also evident in the good agreement between the SO3 and observed EP-flux divergence, which show that most of the planetary wave driving takes place in the upper stratosphere and lower mesosphere (Figure 16). By contrast, the insufficient dissipation of the RF simulated planetary wave in the stratosphere must be compensated by excessive dissipation in the mesosphere. Consequently, the high altitude and low air density of the mesosphere results in the RF simulation being characterised by erroneous strong planetary wave driving in this region (i.e. excessively large EP-flux divergence values).

6. Changes to the representation of tropical winds

Figure 17 compares RF and SO3 zonally averaged winds over the equator (i.e. meridionally averaged between 10°S and 10°N) for 1994 to 2001 with ERAI data. The observations clearly show a QBO signal, with the westerly and easterly components descending down to around the tropopause (50-100 hPa), a phase of around 2 years, and with the westerly component much narrower than the easterly component. The observed SAO is characterised by a higher frequency oscillation of descending westerly and easterly shear zones, with the easterly shear zones descending lower than the westerly shear zones to around 5 hPa.

The SO3 simulation adequately captures the easterly phase of the oscillations, although the easterly QBO winds are too weak. However, the simulation severely underestimates the strength of the westerly QBO winds and fails to capture their required descent, and completely fails to capture the SAO westerly component. Enhancing the launch momentum flux density, $\rho_0 \overline{F}_0^{total}$, at the tropics by both a factor of two and a factor of four using an idealised Gaussian profile, as suggested by Scinocca et al. (2008), did not result in an improvement (not shown). However, the RF representation of tropical winds is still worse: showing a very strong easterly bias, as well a failure to capture the descent of the easterly QBO component and overestimating the frequency and amplitude of the easterly SAO component. The failure of the RF simulation to capture the tropical circulation is a well known problem associated with the use of Rayleigh friction (e.g. Manzini et al., 1997; Scaife et al., 2002). It is also well documented that the implementation of a physically realistic gravity wave drag scheme can lead to an improvement in the representation of tropical winds (e.g. Hamilton et al., 1995; Giorgetta et al., 2002; Manzini et al., 1997; Scaife et al., 2000, 2002; Scinocca et al., 2008). For example, Hamilton et al. (1995) was unable to reproduce the westerly phase of the QBO using a model which resolved planetary waves, but used Rayleigh friction. However, replacing Rayleigh friction with a gravity wave scheme enabled the required westerly momentum flux to be supplied (through gravity wave filtering), resulting in realistic simulation of the westerly phase of these oscillations. Note that both the RF and SO3 simulations show repeating abrupt discontinuities with increasing time. This is a consequence of the simulated 1993 to 2001 data being a composite of eight separate runs, with the abrupt change occurring as the data from one run ends and one run begins.

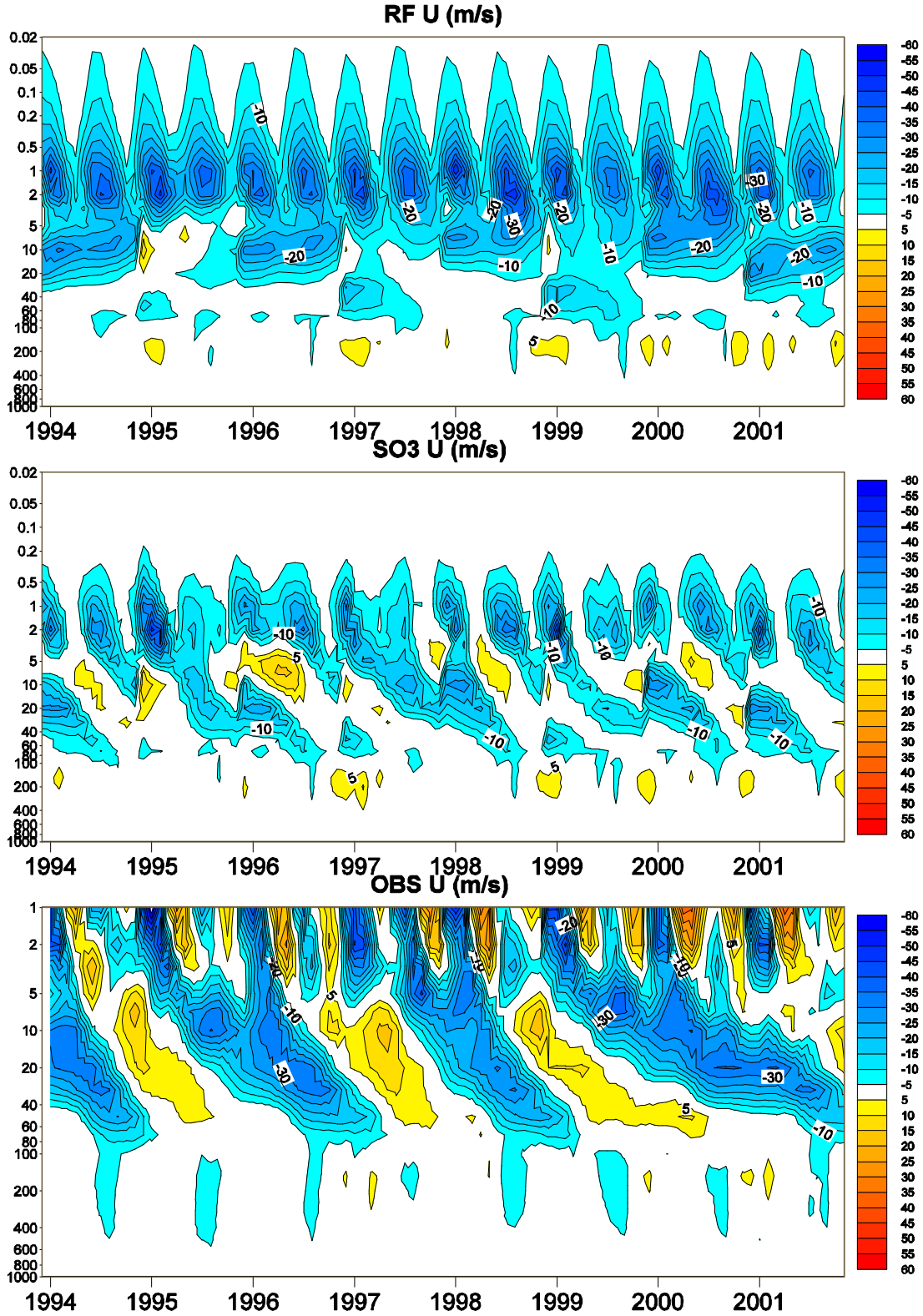

Figure 17: Zonally averaged zonal wind over the equator for the RF (upper) and SO3 (middle) simulations, and observations (lower), for the period 1994 to 2001. The observations are ERAI data. The vertical axes have different scales, with the simulated data stopping at 0.02 hPa and ERAI at 1 hPa. The winds are monthly means and have been meridionally averaged between 10°S and 10°S. The contour interval is 5 m s^{-1}.

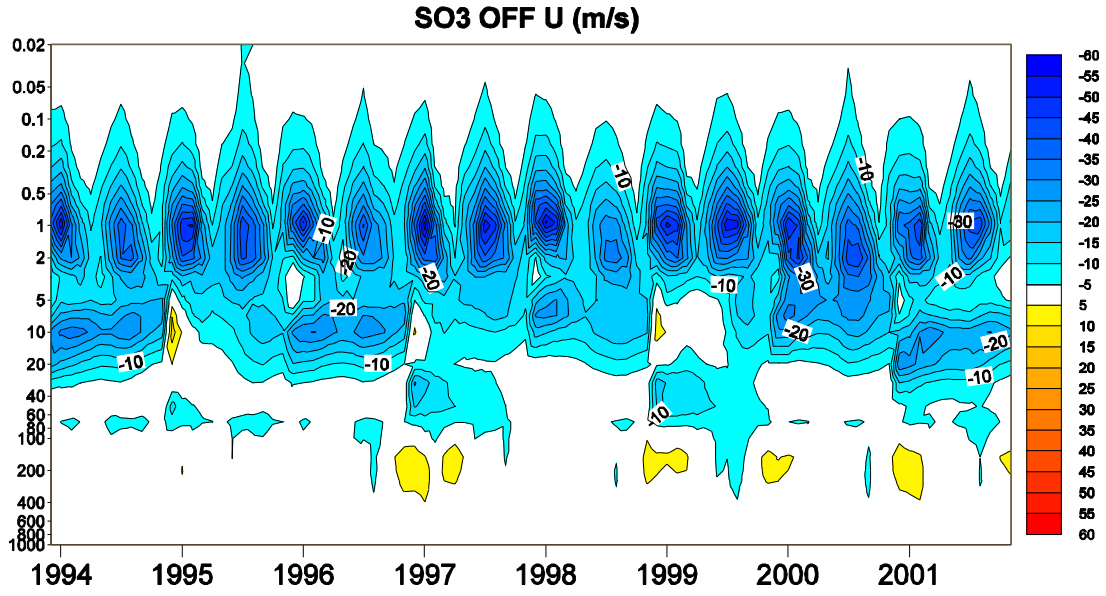

Figure 18: As Figure 17, but showing the results when the SO3 forcing is switched off and the atmosphere is forced by only resolved waves. The vertical axis is pressure (hPa) and extends to 0.02 hPa.

To properly capture tropical oscillations requires a mix of both (resolved) large-scale equatorially trapped Kelvin and Rossby waves, and (unresolved) small-scale gravity waves. The primarily source of the resolved waves is parametrized deep convection in the Tropics (e.g. Ratnam et al., 2004; Scinocca and McFarlane, 2004). Jung et al. (2009) diagnosed Kelvin and Rossby waves by comparing observed and simulated wavenumber-frequency spectra for outgoing longwave radiation, and showed that cycle 33R1 is characterised by a realistic wavenumber-frequency spectra, and that the spectra was correctly dominated by the contribution from tropical convective precipitation, which was much larger than the contribution from large-scale precipitation (Scinocca and McFarlane, 2004). It is perhaps surprising, therefore, that the beneficial impact that this would have had on cycle 33R1 temperatures and winds in the tropical stratosphere and mesosphere, combined with the physically realistic SO3 gravity wave scheme, did not result in the realistic simulation of the tropical winds.

Moreover, switching off the SO3 forcing so as to leave the atmosphere forced only by resolved waves should result in a QBO-like oscillation (J. Scinocca, personal communication). This is presented in Figure 18, but it is readily apparent that any westerly component is completely absent, which perhaps suggests that there is a problem with the resolved waves. This is discussed further in section 10.

7. Comparison of observed and simulated gravity wave momentum fluxes

Ern et al. (2004, 2006) were able to determine the amplitude and the horizontal and vertical wavelengths of stratospheric gravity waves from temperature data obtained on a global scale from satellite missions CRISTA-1 for November 1994 and CRISTA-2 for 8-14 August 1997. These estimates allowed a direct calculation of the absolute value of gravity wave momentum fluxes. Hertzog et al. (2008) showed an excellent agreement between the November 1994 averaged CRISTA-1 momentum fluxes at 22 km and those obtained from long-duration stratospheric balloon observations averaged over November 2005 (the balloon flights were launched between September 2005 and February 2006 over Antarctica and the Southern Ocean as part of the Vorcore campaign).

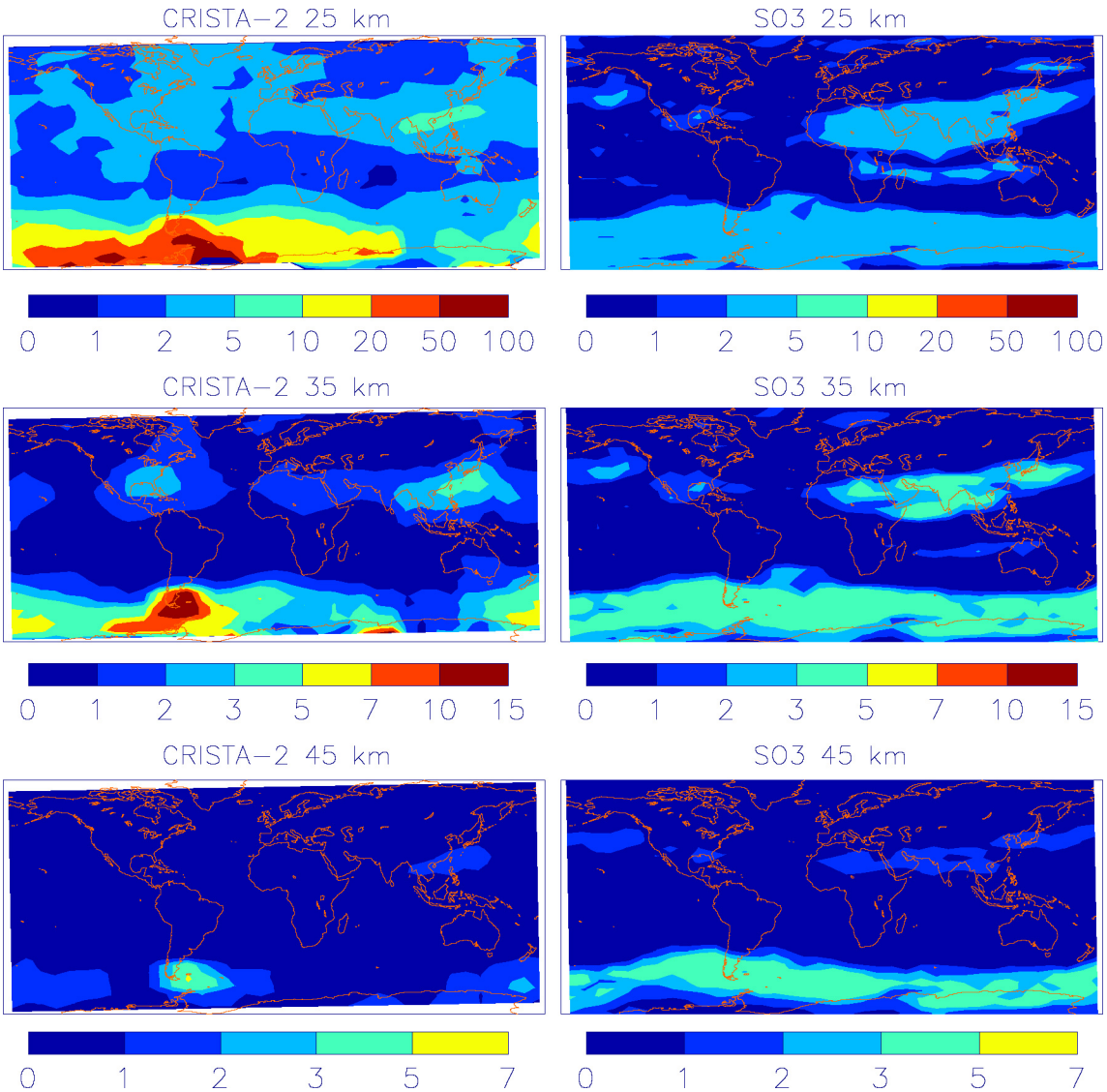

Figure 19: Average CRISTA-2 (left) and SO3 simulated (right) horizontal distribution of absolute values of momentum flux (mPa) for the period 8-14 August 1997 at altitudes of 25 km (upper), 35 km (middle), and 45 km (lower). Contour lines and colour codes are different at different altitudes.

Average CRISTA-2 and SO3 simulated absolute values of gravity wave momentum flux for the period 8-14 August 1997 at altitudes of 25 km, 35 km, and 45 km, are compared in Figure 19. The SO3 fields are computed from the simulation beginning 1 November 1996. At 25 km the observations show a tropical minimum in momentum flux (~1-2 mPa), and localised sub-tropical source regions with enhanced values (~3-6 mPa), such as the Gulf of Mexico and east of China. Enhanced values are also evident at mid and high southern hemisphere latitudes. These features are broadly captured by the SO3 simulation, although the SO3 fluxes are typically a factor of 2-3 smaller than the CRISTA-2 estimates. Moreover, as the SO3 values are not filtered according to the horizontal and vertical wavelengths of gravity waves visible to CRISTA-2, it is likely that such a filtering would reduce their magnitude further. Note that the observed tropical minimum is associated with large temperature variance, with the minimum in momentum flux generated as a consequence of the large horizontal wavelengths (~2000 km) over this region [see Eq. (7) of Ern et al. (2006)]. The very high CRISTA-2 values (~50 mPa) over the Antarctic Peninsula and the southern tip of South America are a consequence of orographic gravity waves, which are not represented by the SO3 momentum flux values. The good agreement in spatial structure is continued at altitudes of 35 and 45 km, with the underestimation of the observed momentum fluxes considerably reduced. Ern et al. (2004) showed a similarly good agreement with momentum fluxes computed by the Warner and McIntyre (2001) parametrization, and suggested that this was a consequence of the (relatively) low altitude of the source spectrum launch height (600 hPa), which allowed the (unrealistic and constant) gravity wave source spectrum to be filtered as it propagated through the realistic tropospheric winds, resulting in a more realistic momentum flux distribution above the tropopause. The low SO3 momentum flux values of the tropical minimum and the localised sub-tropical source regions are an indication that the expected enhancement of gravity wave activity from deep convection in the tropics (e.g. Ratnam et al., 2004; Jiang et al., 2005) must be represented in the SO3 launch spectrum (e.g. Beres et al., 2004, 2005).

8. Sensitivity experiments

This section details the sensitivity of the simulated zonal-mean response to varying the launch spectrum of the SO3 scheme. To achieve this, the 8-year SO3 simulation detailed in section 3 is repeated with differing launch height, total launch momentum flux, and spectral shape.

8.1 Launch height

The average January and July zonal-mean temperature and zonal wind computed using launch heights of 450 and 600 hPa are shown in Figure 20. The remaining parameter settings are $p = 1$, $s = 1$, $m_* = 2\pi / 2000\,\mathrm{m}^{-1}$, $\rho_0 \overline{F}_0^{total} = 3.75 \times 10^{-4}$ Pa, and $n_\phi = 4$. The January northern winter polar temperature response is strikingly different, with the 600 hPa case showing an approximately 10 K cooler/warmer polar stratopause/stratosphere. Examining the zonal wind it is apparent that this is a consequence of the 600 hPa case depositing more momentum in the stratosphere and less momentum in the mesosphere. There could be a number of reasons as to why this is happening (J. Scinocca; personal communication): 1) the critical level filtering will differ as the gravity waves launched from the 600 hPa case must initially travel through a larger range of winds, 2) since the amount of input momentum flux was kept constant, the 450 hPa case will naturally have more momentum at the 450 hPa level, and this difference might be carried on to higher levels, 3) moving the source up and down changes the central phase speed of the launch spectrum, therefore, as the

450 hPa case is in stronger westerlies than the 600 hPa case, it has less easterly momentum (relative to the ground), and 4) more saturation (i.e. momentum deposition) will occur in the upper troposphere/lower stratosphere in the 600 hPa case since its saturation curve (see Figure 1) was defined lower down in the atmosphere. Differences are also apparent in the July southern winter zonal wind, with the 450 hPa case showing a slightly weaker peak velocity of the westerly jet and a more marked tropicward slant.

8.2 Momentum flux

The average January and July zonal-mean response to a total launch momentum flux of 3.5 x 10^{-3} and 4.0 x 10^{-3} Pa is shown in Figure 21. The remaining parameter settings are p = 1, s = 1, $m_* = 2\pi / 2000\text{m}^{-1}$, p_{launch} = 600 hPa, and $n_\phi = 4$. In both the southern and northern winter hemispheres, the peak velocity of the westerly jet decreases and the polar stratosphere temperature increases for the 4.0 x 10^{-3} Pa case. This is consistent with the increased launch momentum flux being associated with more deposition of momentum in the mesosphere and stronger gravity wave drag, with the warmer stratosphere temperature a consequence of the strengthening meridional circulation producing greater downwelling and greater adiabatic warming.

8.3 Source spectrum shape

Figure 22 shows the average January and July zonal-mean response to a small-m spectral slope of 0 and 1. The remaining parameter settings are $p = 1.5$, $m_* = 2\pi / 2000\ \text{m}^{-1}$, p_{launch} = 450 hPa, $\rho_0 \overline{F}_0^{total}$ = 4.25 × 10^{-4} Pa, and $n_\phi = 4$. Unlike in the previous SO3 simulations which employed a value of $p = 1$, here $p = 1.5$. This was necessary as the combination of $(p = 1, s = 0)$ produced an unrealistic circulation. A strong sensitivity is apparent, with the combination of $(p = 1.5, s = 0)$ giving reasonable results. By contrast, the combination of $(p = 1.5, s = 1)$ results in an excessively strong southern winter westerly jet located at a much too high altitude.

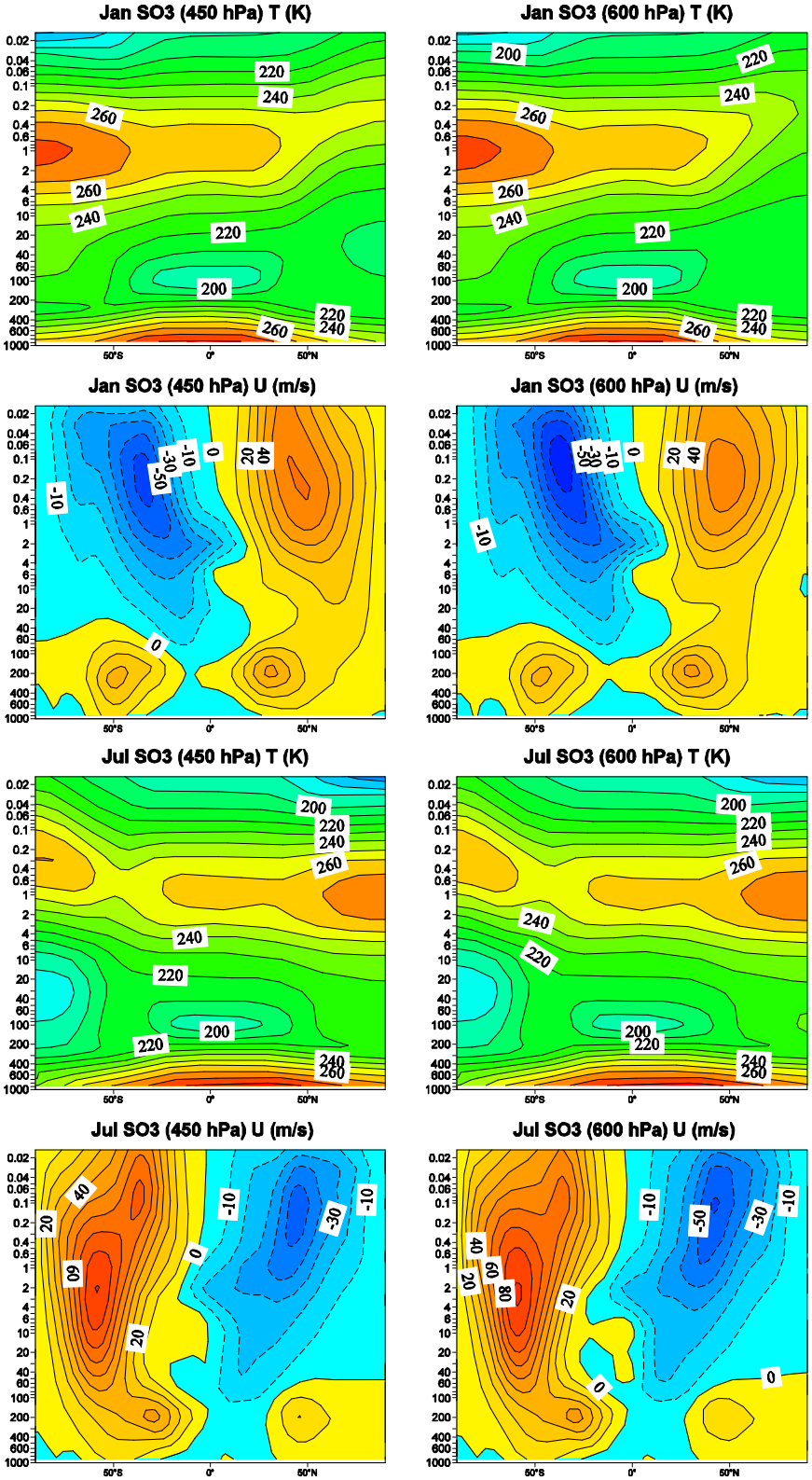

Figure 20: Sensitivity of the SO3 simulated zonal-mean response to launch height for January (upper four panels) and July (lower four panels). Launch heights are 450 hPa (left) and 600 hPa (right). Panels show average cross-sections of zonal-mean temperature and zonal wind. The simulation results are 8-year averages. Contour intervals are 10 K and 10 m s^{-1}. The vertical axis is in pressure (hPa) and extends to 0.015 hPa.

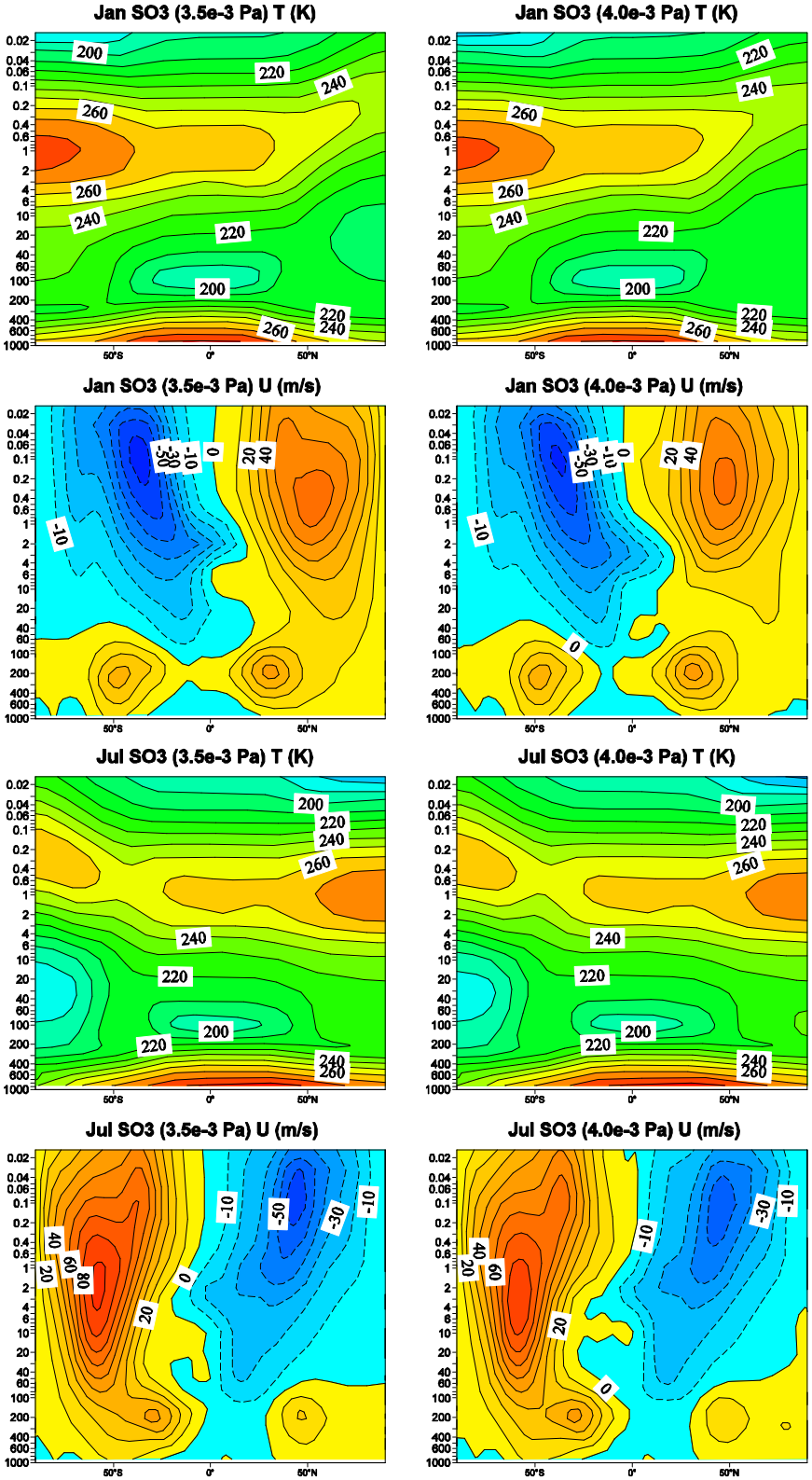

Figure 21: As Figure 20, but showing sensitivity to launch momentum flux. Momentum fluxes are 3.5 x 10^{-3} Pa (left) and 4.0 x 10^{-3} Pa (right).

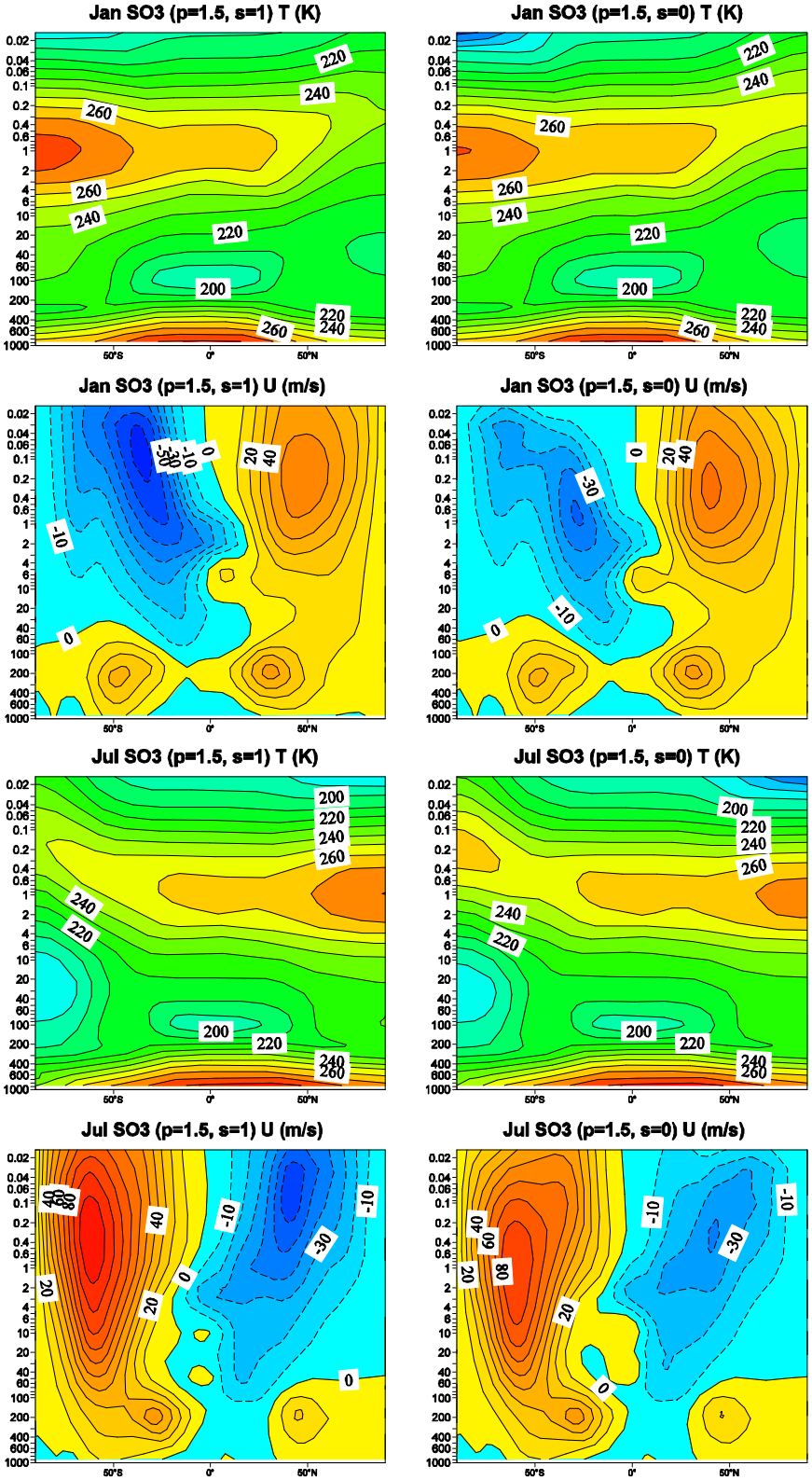

Figure 22: *As Figure 20, but showing sensitivity to the small-m spectral slope, s.* $s = 1$ *(left) and* $s = 0$ *(right). Unlike in Figures 20 and 21 which used a value of* $p = 1$*, here* $p = 1.5$.

9. Evaluation of the Hines Doppler spread parametrization

9.1 Overview of scheme

Like the SO3 scheme, the Hines (1997a, b) Doppler spread parametrization (hereafter DSP) of gravity waves assumes simplified hydrostatic non-rotational wave dynamics. As well as a constant (in space and time) continuous source spectrum, which is isotropic by representing the azimuthal distribution of the direction of wave propagation by discrete azimuths equally spaced around the azimuth circle. Both schemes simplify the problem by integrating out the dependence on $\widetilde{\omega}$ (in the case of SO3) and k (in the case of DSP). However, the SO3 and DSP approaches differ in their treatment of nonlinear dissipation. The DSP scheme assumes that it is wave spectral elements with large-m (i.e. small intrinsic phase speeds) which are most likely to be dissipated, and that as the wave-induced horizontal wind fluctuations become large enough (in response to the diminishing density), they interact nonlinearly with each component of the gravity wave spectrum. Essentially, the wave-induced root-mean-square (rms) fluctuations, U_{rms}, are considered to enhance the background wind, U, in the dispersion relation (3), causing critical level filtering to extend to additional wavenumbers. Following McLandress and Scinocca (2005), the enhanced wind, U_{tot}, is written as

$$U_{tot}(z) = U(z) + U_{rms}(z) = U(z) + \Phi_1 \sigma(z) + \Phi_2 \sigma_T(z) \tag{28}$$

where Φ_1 and Φ_2 are non-dimensional empirical coefficients lying in the range $1.2 < \Phi_1 < 1.9$ and $0.1 < \Phi_2 < 0.4$, σ is the wave-induced rms wind speed fluctuation in the direction of the current azimuth, and σ_T is the total wave-induced rms wind speed wind fluctuation. However, (28) shows that no account is taken of dynamic instability as a consequence of strong vertical wind shear of the background winds (Hines, 1997b), such that the Richardson number, $Ri = N^2/(\partial U/\partial z)^2$, is small and wave breaking is encouraged (Palmer et al., 1986).

9.2 Results

The impact of the DSP scheme was evaluated in a similar fashion to that of the SO3 scheme, with T159L91simulations conducted using cycle 33R1 (+ CLIM_GHG) with the DSP scheme replacing Rayleigh friction. The DSP scheme was implemented with a 'standard' configuration of p_{launch} = 600 hPa, $n_\phi = 8$, $\sigma_{T0} = 0.8$ m s^{-1}, k^* = 2π/(63 km), $\Phi_1 = 1.5$, $\Phi_2 = 0.3$, and $s = 1$, where σ_{T0} is the total wave-induced rms wind at launch height which is split equally among the n_ϕ azimuths, k^* is the effective or aggregate horizontal wavenumber, and s is the slope of the small-m portion of the gravity wave spectrum (as in the SO3 scheme). Fritts and Nastrom (1992) estimated that typical values of σ_{T0} were ~1 m s^{-1}. The simulations comprise eight 13 month integrations covering years 1994 to 2001.

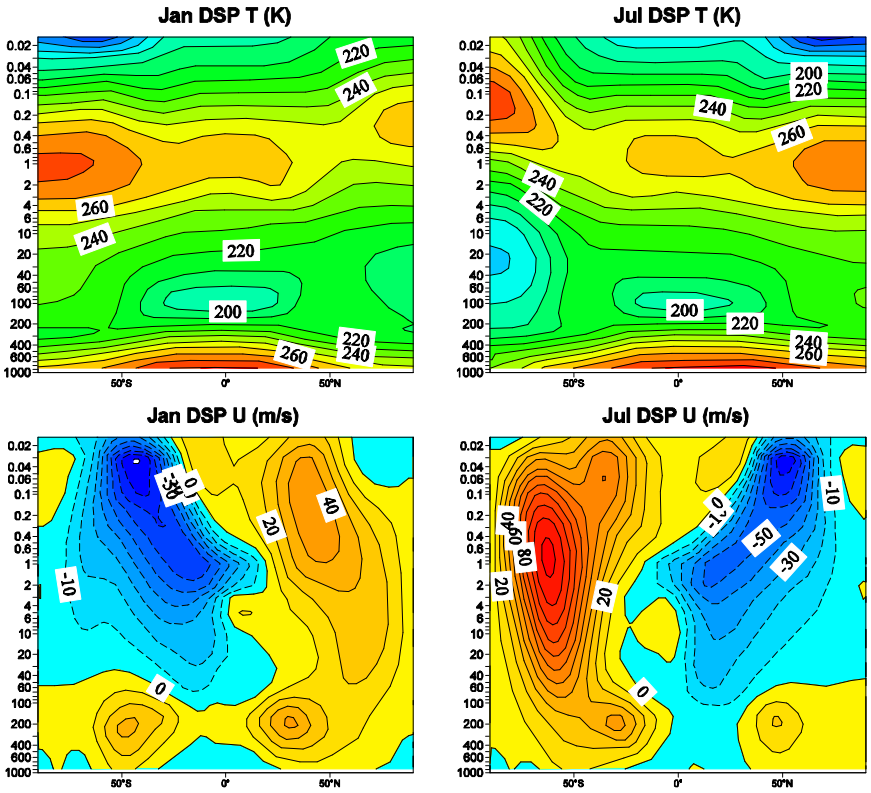

Figure 23: Average January (left) and July (right) cross-sections of zonal-mean temperature (top) and zonal wind (lower) for the DSP simulation. The simulation results are 8-year averages. Contour intervals are 10 K and 10 m s^{-1}. The vertical axis is pressure (hPa) and extends to 0.015 hPa.

The average DSP zonal-mean temperature and zonal wind is displayed in Figure 23 for January and July. The January results suggest that the representation of the northern winter polar stratosphere temperature, northern winter westerlies, and southern summer easterlies, are improved relative to the RF simulation (see Figure 3). However, the northern winter polar stratopause is overly high and warm by around 10 K, suggesting that the required overturning circulation is now excessive as a result of the parametrized gravity wave drag being too strong. Consistent with this is a southern summer polar upper mesosphere temperature that is colder than observed (see Figure 3). This problem is even more exaggerated in the July simulation, with the southern winter polar stratopause temperature reaching around 280 K, as well as being excessively high. Pan et al. (2002) used Fe/Rayleigh lidar measurements and balloon observations to estimate the temperature structure of the middle atmosphere at the southern winter pole, showing that for July that the peak measured stratopause temperature was around 260 K, which is also the value given by the SPARC climatology of Figure 8. Again, consistent with this is a simulated northern summer polar upper mesosphere temperature that is colder than observed. Despite this evidence of strong gravity wave forcing, the resulting strong downwelling has seemingly not reached the southern winter polar stratosphere, which is excessively cold by around 10 K. Linked to this cold bias is an overly strong southern winter westerly jet, with a peak velocity of around 120 m s^{-1}. Furthermore, the jet has none of the expected tropicward tilt, and is characterised by excessive wind shear, which is consistent with the overly warm and high stratopause and of the atmosphere being in thermal balance.

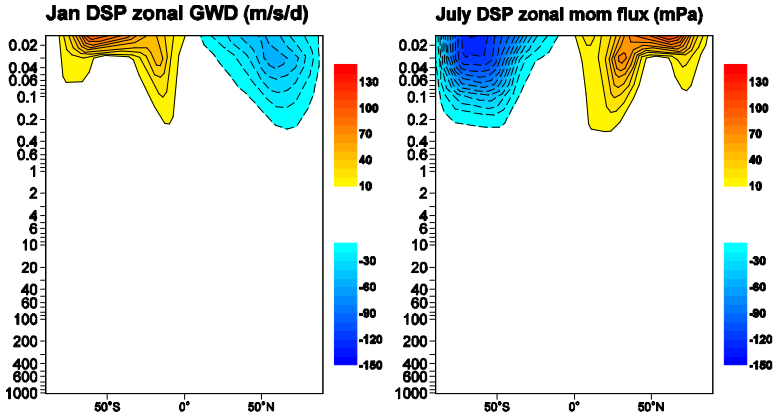

Figure 24: *Average January (left) and July (right) cross-sections of DSP zonal-mean gravity wave drag (m s^{-1} d^{-1}). The simulation results are 8-year averages. The contour interval is ± 10, 20, 30 … 150 m s^{-1} d^{-1}. This differs from that of the equivalent SO3 figures (Figures 6 and 11). The vertical axis is pressure (hPa) and extends to 0.015 hPa.*

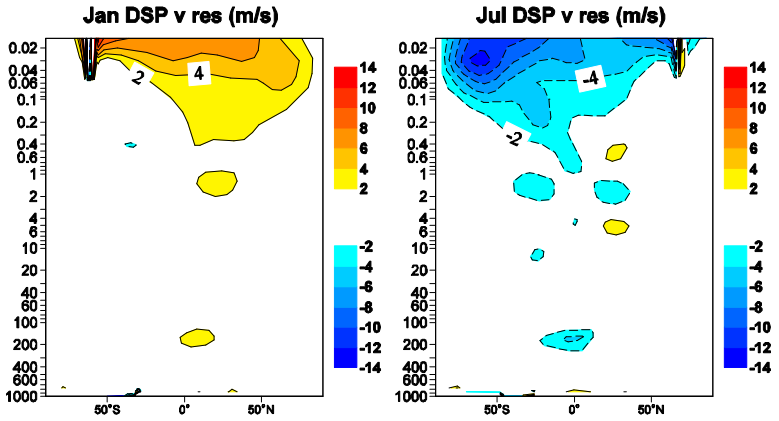

Figure 25: Average January (left) and July (right) cross-sections of DSP zonal-mean residual meridional velocity (m s^{-1}). The simulation results are 8-year averages. The contour interval is ± 2, 4, 6, 8, 10, 12, 14 m s^{-1}. This differs from that of the equivalent SO3 figures (Figures 4 and 9). The vertical axis is pressure (hPa) and extends to 0.015 hPa.

Figure 24 shows that the average DSP zonal-mean gravity wave drag for January and July are 2-3 times greater than the equivalent SO3 values (see Figures 6 and 11), with winter values peaking at -150 and -50 m s^{-1} d^{-1} in the southern and northern hemispheres respectively. These estimates are also several times larger than the observed peak winter values previously stated in section 4, which were -50 and -20 m s^{-1} d^{-1} in the southern and northern hemispheres respectively. Additionally, in contrast to the SO3 scheme, the DSP drag is mostly strongly confined over the uppermost few levels of the model, i.e. at the top of the upper mesosphere. This is also in disagreement with the available observations, which showed that the peak drag should be located in the lower mesosphere (Smith and Lyjak, 1985; Pulido and Thuburn, 2006). But is consistent with McLandress and Scinocca (2005) who showed that for a similar gravity wave source that the momentum flux computed by the DSP scheme decreased at a slower rate than that of the SO3 scheme, i.e. the DSP scheme deposits its momentum higher in the atmosphere than the SO3 scheme, which from (21) results in much larger drag values.

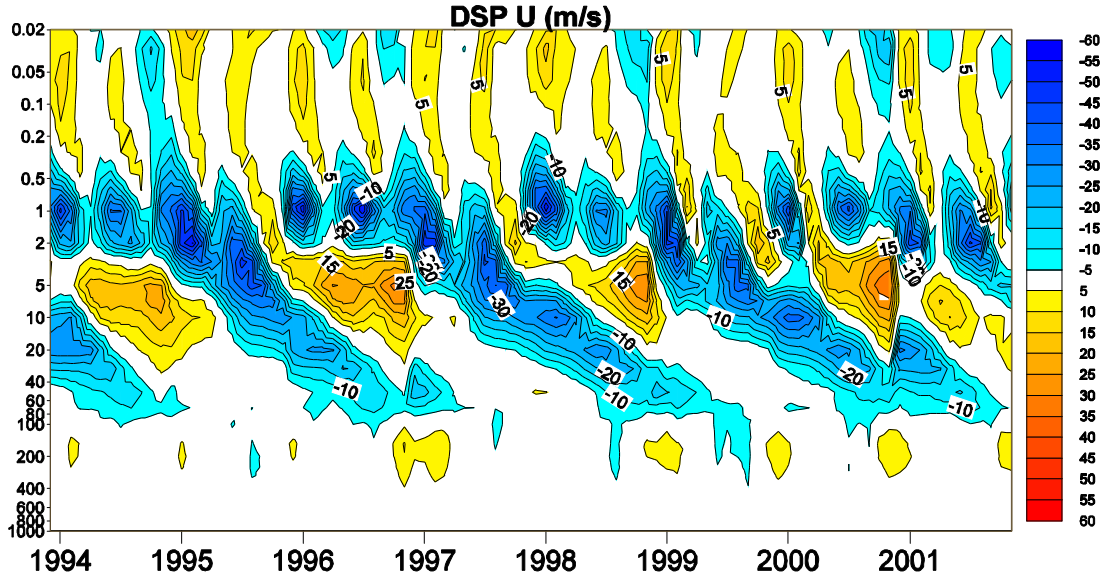

Figure 26: As Figure 17, but showing the zonally averaged zonal wind over the equator for the DSP simulation. The vertical axis is pressure (hPa) and extends to 0.02 hPa.

This is perhaps a consequence of the DSP scheme not including the contribution to dynamic instability of the background wind shear, which is particularly considerable in the lower and middle mesosphere regions of the southern winter westerly jet, and which are known to be preferential wave breaking regions. The DSP scheme was subsequently modified to additionally include the Richardson-number-dependent gravity wave breaking formulation of Palmer et al. (1986), which considers the influence of static stability and background vertical wind shear on wave breaking. Surprisingly, perhaps, this did not result in any significant improvement of the zonal-mean temperature and zonal wind (not shown), with an excessively high and warm winter stratopause still apparent, suggesting that the methodology applied by the DSP scheme to capture the process of nonlinear dissipation is fundamentally flawed. Furthermore, Hines (1997a, b) encourages the use of $n_\phi = 16$ azimuths of propagation so as to capture as well as possible the interaction between waves. However, the atmospheric response was relatively insensitive to this change (not shown). Figure 25 shows that average January and July values of the zonal-mean residual meridional velocity are around 1.5 and 2 times greater, respectively, than the equivalent SO3 values (Figures 4 and 9), confirming that the DSP simulation is characterised by an overly strong overturning circulation.

By contrast, the DSP simulation results in a marked improvement in the representation of the QBO and SAO (Figure 26). The DSP simulation adequately captures the phase of the QBO signal evident in the observations, although the westerly winds are too strong and only descend down as far as 20 hPa. Regarding the SAO component, the DSP simulation captures the phase, amplitude, and altitude of the easterly component. However, the westerly shear component is absent, which is perhaps a result of the continued need to apply Rayleigh friction over a narrow region of the tropics. Scaife et al. (2002) similarly show that the westerly phase of the QBO simulated by the MetOffice Unified Model does not descend far enough. They suggested that, particularly towards the end of its descent where it narrows, that the westerly phase is being directly damped by the model horizontal diffusion, and showed that reducing the horizontal diffusion resulted in improved descent.

10. Discussion and conclusion

The SO3 gravity wave drag parametrization describes the vertical evolution of a broad, constant, isotropic spectrum of gravity waves, emanating from the troposphere. The scheme models critical level filtering, conservative propagation, and nonlinear dissipation, which is dealt with in an empirical fashion by limiting the growth of the gravity wave spectrum to not exceed some saturated spectrum proportional to m^{-3}. Replacing Rayleigh friction with the SO3 scheme in cycle 33R1 (+ CLIM_GHG) showed an improvement in the qualitative agreement between the IFS model climate and observations, resulting in:

i) Increased dynamical warming of the southern winter pole, resulting in an alleviation of the southern winter polar stratosphere cold bias.

ii) A strong reduction in the westerly bias of the southern winter westerly jet, along with improved tropicward tilt, resulting in realistic southern winter zonal-mean winds.

iii) Stronger and more realistic summer easterly jets in both the northern and southern summer hemispheres.

iv) An alleviation of the summer polar upper mesosphere warm bias and a realistic summer to winter hemisphere meridional circulation, resulting in a stronger and more realistic overturning circulation.

v) Good agreement between parametrized and observed summer and winter hemisphere values of gravity wave drag.

vi) Good agreement between parametrized and absolute values of gravity wave momentum flux derived from satellite data.

vii) A more realistic reduction in southern winter orographic gravity wave momentum deposition at the model top due to earlier (lower down) orographic wave breaking as a consequence of the improved closure of the southern winter westerly jet.

viii) Good agreement between simulated and observed stationary planetary wave amplitudes in the stratosphere, resulting in realistic wave-mean flow driving of the stratosphere and mesosphere.

Unfortunately, operational ensemble monthly and seasonal forecasts cannot use the SO3 scheme as they use a 62 level configuration with a model top at 5 hPa, which does not resolve the mesosphere.

However, a notable failure of the SO3 simulation was the inability to capture the variability of tropical stratospheric circulation, i.e. the QBO and the SAO, despite Scinocca et al. (2008) having achieved this with the SO3 scheme using the Canadian Centre for Climate Modelling and Analysis (CCCma) GCM. Switching off the SO3 forcing so as to leave the tropical atmosphere forced only by resolved large-scale equatorially trapped Kelvin and Rossby waves did not result in a QBO-like oscillation, which is suggestive of a problem with the resolved wave forcing. However 1) the Doppler spread parametrization (DSP) simulation realistically captured the QBO, and 2) Jung et al. (2009) showed that cycle 33R1 was characterised by realistic equatorially trapped resolved waves, which were correctly dominated by the contribution from tropical convective precipitation (Scinocca and McFarlane, 2004). It is possible that the failure to capture a

QBO-like oscillation with just resolved forcing was because of damping by the Rayleigh friction on the zonal-mean flow, which remained on in the SO3 simulation over a narrow region at the tropics [see Eq. (27)]. This might also damp the contribution from the SO3 forcing in the simulation with both parametrized and resolved forcing. While the DSP simulation was able to capture the QBO as it is characterised by much stronger gravity wave forcing relative to the SO3 scheme [see section 9 and McLandress and Scinocca (2005)], which was not completely negated by Rayleigh friction. This issue warrants further attention.

It is also possible that the failure to capture a QBO-like oscillation with just resolved forcing is because of the relatively crude vertical resolution in the tropical stratosphere and mesosphere. As stated, this is already the cause of a spurious stratospheric jet in the tropics which required to be damped by Rayleigh friction. Hamilton et al. (1999) performed integrations with a GCM which only explicitly resolved waves (i.e. no parametrized gravity wave drag), showing that the simulated tropical stratospheric winds were extremely dependent on vertical resolution. With a lower stratospheric vertical resolution of ~1.5 km they showed that the zonal-mean tropical winds were nearly constant in time. But by doubling the vertical resolution they showed the formation of a QBO-like oscillation. Scinocca et al. (2008) similarly demonstrated a strong sensitivity to vertical resolution using the CCCma model. The 91 vertical levels of the IFS have a vertical resolution of around ~1 km in the tropical stratosphere, which these arguments suggest is too coarse. This conclusion is consistent with the study of Boville and Randel (1992) who estimated that a vertical resolution of about 1 km or less was required to represent equatorially-trapped waves adequately. Improved vertical resolution within the stratosphere and mesosphere would therefore be of benefit. Improved vertical resolution would also allow better representation of breaking gravity waves in the mesosphere and of the vertical distribution of the parametrized gravity wave drag.

One of the difficulties of the SO3 scheme is the broad range of tuning parameters required to define its gravity wave launch spectrum, and which are only loosely constrained by limited observations. Furthermore, limited observations of dissipating gravity waves in the mesosphere require that the SO3 scheme is effectively tuned to simulate the observed climate. However, considerable advances have been made recently in the observations of gravity waves, providing stronger observational constraints on the gravity wave characteristics and perhaps reducing the need for extensive tuning. For example, Alexander and Rosenlof (2003) used Upper Atmosphere Research Satellite (UARS) data to derive estimates of the gravity wave mean flow forcing in the stratosphere, and from this inferred constraints on gravity wave characteristics near the tropopause, showing that there are substantial differences between waves in the tropics and waves in the extratropics. Under this constraint, it is immediately apparent that a globally uniform launch spectrum is unsuitable, and that the introduction of a latitudinal varying source spectrum to account for the pronounced peak in gravity wave activity at tropical latitudes would be of benefit. Scaife et al. (2000) adopted this method after first choosing a launch spectrum which produced a realistic QBO, i.e. accepting that the QBO is forced by a sizeable contribution from gravity waves (Lindzen and Holton, 1968; Alexander and Holton, 1997). Yang et al. (2006) adopted separate input sources for the tropics and extratropics. Alexander and Rosenlof (2003) suggested that a seasonally varying input source is required. Alexander et al. (2002) argued that the peak in gravity wave activity at tropical latitudes may be a natural consequence of the latitudinal variation in the Coriolis parameter, which controls the lower limit of the gravity wave intrinsic frequency.

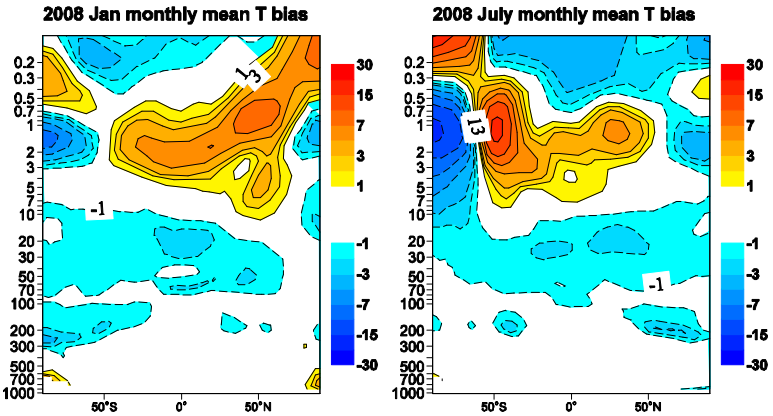

Figure 27: Average T+240 h monthly temperature bias (K) at operational T799 resolution for January (left) and July (right) 2008. Vertical axis is pressure (hPa) and extends to 0.1 hPa.

Alternatively, to better capture more realistic spatial and temporal distributions of gravity wave activity, some direct dependence of the source spectrum to localised sources of strong gravity wave excitation such as convection, fronts, and the jet-stream could be introduced. For example, Beres et al. (2004, 2005) implemented a gravity wave source spectrum dependent on the properties of convection such as the convective heating rate and depth, and background wind, leading to an improved representation of the SAO. Rind et al. (1988) included a parametrization of gravity wave drag contributions from flow over orography, wind shear, and convection, in the GISS global climate model, and showed generally realistic temperature and wind fields throughout the stratosphere and mesosphere up to approximately 75 km, including a realistic closing off of the winter jet in the mesosphere and winter polar mesosphere temperatures. While Charron and Manzini (2002) modified the gravity wave launch spectrum to represent the enhanced gravity wave wind variance and direction of propagation associated with the arrival of a frontal system (apparent by an increase in horizontal potential temperature gradient).

Finally, studies such as Hamilton et al. (1995, 1999) show a convergence of results with increasing horizontal resolution, showing with no explicit representation of gravity wave drag a southern winter polar stratosphere bias of ~70 K at a horizontal resolution of ~300 km and a bias of ~10 K at ~35 km resolution. This suggests there is still a large contribution from 'missing' gravity waves at a resolution of ~35 km which require parametrization. Even at this relatively high resolution it is apparent that the drag supplied by Rayleigh friction is insufficient. This is readily apparent in the IFS operational forecast, which currently has a horizontal resolution of around 25 km. Figure 27 shows the average T + 240 h operational forecast temperature bias for January (cycle 32R3) and July (cycle 33R1), 2008, showing a cold temperature bias of ~20 K at the southern winter polar stratosphere, consistent with an overly weak overturning circulation and a lack of dynamical warming. Operational implementation of the SO3 scheme might therefore result in a reduction in biases. Tropospheric predictions may be improved by the realistic overturning circulation which affects polar temperatures down as far as the lower stratosphere and the more realistic variability in tropical stratospheric winds. The impact of the SO3 scheme on short and medium range deterministic forecasts requires further examination. Moreover, the issue of 'double counting' might arise at higher horizontal resolutions; in that the parametrized gravity wave drag begins to account for gravity waves which are now resolved by the model. Currently there is no explicit sensitivity to horizontal resolution in the SO3 scheme, such as decreasing the launch momentum flux with increasing horizontal resolution. Although, the model will partly regulate itself by simulating typically weaker and more realistic background winds with

increasing horizontal resolution (i.e. as more of the wave spectrum is resolved), which would filter the gravity wave spectrum as it propagates through the stratosphere so that the waves carry less momentum flux as they enter the mesosphere. Thus when they are dissipated the forcing is small. Resolution dependence of the scheme may also be aided by linking the source spectrum to sub-grid precipitation.

Acknowledgements

J. Scinocca at the Canadian Centre for Climate Modelling and Analysis, University of Victoria, and E. Manzini at the Istituto Nazionale de Geofisica e Vulcanologia, Bologna, respectively supplied the SO3 and DSP gravity wave drag parameterizations, as well as many valuable comments and advice. A. Beljaars at ECMWF provided helpful comments on a draft version of this memorandum. T. Jung at ECMWF provided help and code to analyse the EP flux divergence and in running the climate simulations. The improved GHG climatology was implemented in the IFS by J.-J. Morcrette at ECMWF. M. Ern at Forschungszentrum Jülich, Germany provided the CRISTA-2 satellite data. Helpful conversations were had with P. Bechtold, P. Bougeault, P. Jansen, T. Jung, M. Miller, J.-J. Morcrette, A. Untch, and N. Wedi at ECMWF, as well as A. Bushell and D. Jackson at the MetOffice, D. Cariolle at Cerfacs, M. Giorgetta at the Max Planck Institute, Hamburg, F. Lott at Laboratoire de Meteorologie Dynamique, IPSL, and D. Saint-Martin at MeteoFrance. Much kind assistance from F. Ii and S. Lamy-Thepaut at ECMWF on producing the figures is gratefully acknowledged.

References

Akmaev, R. A., 'Simulation of large-scale dynamics in the mesosphere and lower thermosphere with the Doppler-spread parameterization of gravity waves. 1. Implementation and zonal mean climatologies', *J. Geophys. Res.*, **106**, pp. 1193-1204, 2001.

Alexander, M. J., and J. R. Holton, 'A model study of the zonal forcing in the equatorial stratosphere by convectively induced gravity waves', *J. Atmos. Sci.,* **54**, 1408-1419, 1997.

Alexander, M. J., and T. J. Dunkerton, 'A spectral parameterization of mean-flow forcing due to breaking gravity waves', *J. Atmos. Sci.*, **56**, pp. 4167-4182, 1999.

Alexander, M. J., and K. H. Rosenlof, 'Gravity-wave forcing in the stratosphere: Observational constraints from the Upper Atmosphere Research Satellite and implications for parameterization in global models', *J. Geophys. Res*., **108**, 4597, doi:10.1029/2003JD003373, 2003.

Alexander, M. J., T. Tsuda, and R. A. Vincent, 'Latitudinal variations observed in gravity waves with short vertical wavelengths', *J. Atmos. Phys.*, **59**, pp. 1394-1404, 2002.

Allen, S. J., and R. A. Vincent, 'Gravity wave activity in the lower atmosphere: Seasonal and latitudinal variations', *J. Geophys. Res.*, **100**, pp. 1327-1350, 1995.

Andrews, D. G., J. R. Holton, and C. B Leovy, *Middle Atmosphere Dynamics*, International Geophysical Series, Vol. 40, pp. 489, 1987.

Beres, J. H., M. J. Alexander, and J. R. Holton, 'A method of specifying the gravity wave spectrum above convection based on latent heating properties and background wind', *J. Atmos. Sci.*, **61**, pp. 324-337, 2004.

Beres, J. H., R. R. Garcia, B. A. Boville, and F. Sassi, 'Implementation of a gravity wave source spectrum parameterization dependent on the properties of convection in the Whole Atmosphere Community Climate Model (WACCM)', *J. Geophys. Res.*, **110**, D10108, doi:10.1029/2004JD005504, 2005.

Boville, B., A., 'Wave mean flow interactions in a general circulation model of the troposphere and stratosphere', *J. Atmos. Sci.*, **43**, pp. 1711-1144, 1986.

Boville, B. A., and W. J. Randel, 'Equatorial waves in a stratospheric GCM: effects of vertical resolution', *J. Atmos. Sci.*, **49**, pp. 785-801, 1992.

Cariolle, D., M. J. Evans, M. P. Chipperfield, N. Butkovskaya, A. Kukui, and G. Le Bras, 'Impact of the new HNO_3-forming channel of the HO_2+NO reaction on tropospheric HNO_3, NO_x, HO_x and ozone', *Atmos. Chem. Phys*., **8**, pp. 4061-4068, 2008.

Charron, M., and E. Manzini, 'Gravity waves from fronts: parameterization and middle atmosphere response in a general circulation model', *J. Atmos. Sci.*, **59**, pp. 923-941, 2002.

Dewan, E. M, and R. E. Good, 'Saturation and the "universal" spectrum for vertical profiles of horizontal scalar winds in the atmosphere', *J. Geophys. Res.*, **91**, pp. 2742-2748, 1986.

Eckermann, S. D, and R. A. Vincent, 'VHF radar observations of gravity-wave production by cold fronts over Southern Australia', *J. Atmos. Sci.*, **50**, pp. 785-806, 1993.

Eckermann, S. D., and P. Preusse, 'Global measurements of stratospheric mountain waves from space', Science, **286**, pp. 1534-1537, 1999.

Edmon, H. J., B. J. Hoskins, and M. E. McIntyre, 'Eliassen-Palm cross sections for the troposphere', *J. Atmos. Sci.*, **37**, pp. 2600-2616, 1980.

Ern, M., P. Preusse, M. J. Alexander, and C. D. Warner, 'Absolute values of gravity wave momentum flux derived from satellite data', *J. Geophys. Res*., **109**, D20103, doi:10.1029/2004JD004752, 2004.

Ern, M. P., P. Preusse, and C. D. Warner, 'Some experimental constraints for spectral parameters used in the Warner and McIntyre gravity wave parameterization scheme', Atmos. Chem. Phys., **6**, pp. 4361-4381, 2006.

Fomichev, V. I., W. E. Ward, S. R. Beagley, C. McLandress, J. C. McConnell, N. A. McFarlane, and T. G. Shepherd, 'The extended Canadian Middle Atmosphere Model: zonal-mean climatology and parmeterizations', *J. Geophys. Res.*, **107**, doi:10.1029/2001JD000479, 2002.

Fritts, D. C., and G. D. Nastrom, 'Sources of mesoscale variability of gravity waves. Part II: frontal, convective, and jet stream excitation', *J. Atmos. Sci.*, **49**, pp. 111-127, 1992.

Fritts, D. C., and T. E. VanZandt, 'Spectral estimates of gravity wave energy and momentum fluxes. Part I: Energy dissipation, acceleration, and constraints. *J. Atmos. Sci.*, **50**, pp. 3685-3694, 1993.

Fritts, D. C, and M. J. Alexander, 'Gravity wave dynamics and effects in the middle atmosphere', *Rev. Geophys.*, **41(1)**, 1003, doi:10.1029/2001RG000106, 2003.

Garcia, R. R., and S. Solomon, 'The effect of breaking gravity waves on the dynamics and chemical composition of the mesosphere and lower thermosphere', *J. Geo. Res.*, **90,** pp. 3850-3868, 1985.

Garcia, R. R., and B. A. Boville, ' "Downward Control" of the mean meridional circulation and temperature distribution of the polar winter stratosphere', *J. Atmos. Sci.*, **51**, pp. 2238-2245, 1994.

Gill, A. E., *Atmosphere-ocean dynamics*, Academic Press, 662 pp., 1982.

Giorgetta, M. A., E. Manzini, and E. Roeckner, 'Forcing of the quasi-biennial oscillation from a broad spectrum of atmospheric waves', *Geophys. Res. Lett.*, **29**, doi: 10.1029/2002GL014756, 2002.

Hamilton, K., R. J. Wilson, J. D. Mahlman, and L. J. Umscheid, 'Climatology of the SKYHI troposphere-stratosphere-mesosphere general circulation model', *J. Atmos. Sci.*, **52**, pp. 5-42, 1995.

Hamilton, K., R. J. Wilson, and R. S. Hemler, 'Middle atmosphere simulated with high vertical and horizontal resolution versions of a GCM: Improvements in the cold pole bias and generation of a QBO-like oscillation in the tropics', *J. Atmos. Sci.*, **56**, pp. 3829-3846, 1999.

Haynes, P. H., C. J. Marks, M. E. McIntyre, T. G. Shepherd, and K. P. Shine, 'On the "downward control" of extratropical diabatic circulations by eddy-induced mean forces', *J. Atmos. Sci.*, **48**, pp. 651-678, 1991.

Hertzog, A., G. Boccara, R. A. Vincent, F. Vial, and P. Cocquerez, 'Estimation of gravity wave momentum flux and phase speeds from quasi-lagrangian stratospheric balloon flights. Part II: Results from the Vorcore campaign in Antarctica', *J. Atmos. Sci.*, **65**, pp. 3056-3070, 2008.

Hines, C. O., 'Doppler spread parameterization of gravity wave momentum deposition in the middle atmosphere. Part 1: Basic formulation', *J. Atmos., Solar Terr., Phys.*, **59**, pp. 371-386, 1997a.

Hines, C. O., 'Doppler spread parameterization of gravity wave momentum deposition in the middle atmosphere. Part 2: Broad and quasi monochromatic spectra and implementation', *J. Atmos., Solar Terr., Phys.*, **59**, pp. 387-400, 1997b.

Hitchman, M. H., and C. B. Leovy, 'Evolution of the zonal mean state in the equatorial middle atmosphere during October 1978 - May 1979', *J. Atmos. Sci.*, **43**, pp. 3159-3176, 1986.

Holton, J. R., 'The role of gravity wave induced drag and diffusion in the momentum budget of the mesosphere', *J. Atmos. Sci.*, **39**, pp. 791-799, 1982.

Holton, J. R., 'The influence of gravity wave breaking on the general circulation of the middle atmosphere', *J. Atmos. Sci.*, **40**, pp. 2497-2507, 1983.

Holton, J. R., *An introduction to dynamic meteorology*, Elsevier/Academic Press, fourth edition, 535 pp., 2004.

Jiang, J. H., S. D. Eckermann, D. L. Wu, K. Hocke, B. Wang, J. Ma, and Y. Zhang, 'Seasonal variation of gravity wave sources from satellite observations', *Adv. Space Res.*, **35**, pp. 1925-1932, 2005.

Jung, T., G. Balsamo, P. Bechtold, A. Beljaars, M. Kohler, M. Miller, J.-J. Morcrette, A. Orr, M. Rodwell, A. Tompkins, A. Untch, and N. Wedi, 'The ECMWF model climate: recent progress through improved physical parametrizations', *Q. J. R. Met. Soc*, submitted, 2009.

Lane, T. P., J. D. Doyle, R. Plougonven, M. A. Shapiro, and R. D. Sharman, 'Observations and numerical simulations of inertia-gravity waves and shearing instabilities in the vicinity of a jet stream', *J. Atmos. Sci.*, **61**, pp. 2692-2706, 2008.

Lawrence, B. N., 'Some aspects of the sensitivity of stratospheric climate simulation to model lid height', *J. Geophys. Res.*, **102**, pp. 23,805-23,811, 1997.

Lindzen, R. S., and J. R. Holton, 'Theory of the stratospheric quasi-biennial oscillation', *J. Atmos. Sci.*, **25**, pp. 1095-1107, 1968.

Lott, F. L., and M. Miller, 'A new subgrid scale orographic drag parametrization; its formulation and testing', *Q. J. R. Met. Soc.*, **123**, pp. 101-127, 1997.

Matsuno, T., and K. Nakamura, 'The eulerian- and lagrangian-mean meridional circulations in the stratosphere at the time of a sudden warming', *J. Atmos. Phys*., **36**, pp. 640-654, 1979.

Manzini, E., N. A. McFarlane, and C. McLandress, 'Impact of the Doppler spread parameterization on the simulation of the middle atmosphere circulation using the MA/ECHAM4 general circulation model', *J. Geophys. Res.*, **102**, pp. 25751-25762, 1997.

Manzini, E., and N. A. McFarlane, 'The effect of varying the source spectrum of a gravity wave parameterization in a middle atmosphere general circulation model', *J. Geophys. Res.*, **103**, pp. 31523-31539, 1998.

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

McLandress, C., and J. F. Scinocca, 'The GCM response to current parameterizations of nonorographic gravity wave drag', *J. Atmos. Sci.*, **62**, pp. 2394-2413, 2005.

McNally, T., 'The assimilation of stratospheric satellite data at ECMWF', ECMWF/SPARC workshop on modelling and assimilation for the stratosphere and tropopause', 23 to 26 June, 2003.

Medvedev, A. S., and G. P. Klaassen, 'Vertical evolution of gravity wave spectra and the parameterization of associated wave drag', *J. Geophys. Res.*, **100**, pp. 25841-25853, 1995.

Medvedev, A. S., G. P. Klaassen, and S. R. Beagley, 'On the role of an anisotropic gravity wave spectrum in maintaining the circulation of the middle atmosphere', *Geophys. Res. Lett.*, **25**, pp. 509-512, 1998.

Morcrette, J.-J., and co-authors, 'Recent advances in radiation transfer parameterizations', ECMWF Technical Memorandum, No. 539, 2007.

Norton, W. A., and J. Thuburn, 'Sensitivity of mesospheric mean flow, planetary waves, and tides to strength of gravity wave drag', *J. Geophys. Res.*, **104**, pp. 30,897-30,911, 1999.

Orr, A., and N. Wedi, 'The representation of non-orographic gravity waves in the IFS. Part I: Assessment of the middle atmosphere model climate with Rayleigh friction', ECMWF technical memo, No. 592, 2009.

Palmer, T. N., G. J. Shutts, and R. Swinbank, 'Alleviation of a systematic westerly bias in general circulation and numerical weather prediction models through an orographic gravity wave drag parameterization', *Q. J. R. Met. Soc.*, **112**, pp. 1001-1039, 1986.

Pan, W., C. S. Gardner, and R. G. Roble, 'The temperature structure of the winter atmosphere at South Pole', *Geophys. Res. Lett.*, **29**, doi:10.1029/2002GL015288, 2002.

Pulido, M., and J. Thuburn, 'Gravity-wave drag estimation from global analyses using variational data assimilation principles. II: Case study', *Q. J. Roy. Met. Soc.*, **132**, pp. 1527-1543, 2006.

Randel, W. J., and co-authors, 'The SPARC intercomparison of middle atmosphere climatologies', *J. Climate*, **17**, pp. 986-1003, 2004.

Ratnam, M. V., G. Tetzlaff, and C. Jacobi, 'Global and seasonal variations of stratospheric gravity wave activity deduced from the CHAMP/GPS satellite', *J. Atmos. Sci.*, **61**, pp. 1610-1620, 2004.

Rind, D., R. Suozzo, N. K. Balachandran, A. Lacis, and G. Russell, 'The GISS global climate-middle atmosphere model. Part I: model structure and climatology', *J. Atmos. Sci*., **45**, pp. 329-369, 1988.

Scaife, A. A., N. Butchart, C. D. Warner, D. Stainforth, W. Norton, and J. Austin, 'Realistic quasi-biennial oscillations in a simulation of the global climate', *Geo. Res. Lett.*, **27**, pp. 3481-3484, 2000.

Scaife, A. A., N. Butchart, C. D. Warner, and R. Swinbank, 'Impact of a spectral gravity wave parameterization on the stratosphere in the Met Office Unified Model', *J. Atmos. Sci.*, **59**, pp. 1473-1489, 2002.

Scinocca, J. F., 'The effect of back-reflection in the parameterization of non-orographic gravity-wave drag', *J. Met. Soc. Jap.*, **80**, pp. 939-962, 2002.

Scinocca, J. F., 'An accurate spectral nonorographic gravity wave drag parameterization for general circulation models', *J. Atmos. Sci.*, **60**, pp. 667-682, 2003.

Scinocca, J. F., and N. A. McFarlane, 'The variability of modeled tropical precipitation', *J. Atmos. Sci.*, **61**, pp. 1993-2015, 2004.

Scinocca, J. F., N. A. McFarlane, M. Lazare, J. Li, and D. Plummer, 'The CCCma third generation AGCM and its extension into the middle atmosphere', *Atmos. Chem. Phys. Discuss.*, **8**, pp. 7883-7930, 2008.

Shaw, T. A., and T. G. Shepherd, 'Angular momentum conservation and gravity wave drag parameterization: implications for climate models', *J. Atmos. Sci.*, **64**, pp. 190-203, 2007.

Shepherd, T. G., K. Semeniuk, and J. N. Koshyk, 'Sponge layer feedbacks in middle-atmosphere models', *J. Geophys. Res.*, **101**, pp. 23,447-23,464, 1996.

Shepherd, T. G., and T. A. Shaw, 'The angular momentum constraint on climate sensitivity and downward influence in the middle atmosphere', *J. Atmos. Sci.*, **61**, pp. 2899-2908, 2004.

Shine, K., 'Sources and sinks of zonal momentum in the middle atmosphere diagnosed using the diabatic circulation', *Q. J. R. Met. Soc.*, **115**, pp. 265-292, 1989.

Smith, S. A., D. C. Fritts, and T. E. VanZandt, 'Evidence for a saturated spectrum of atmospheric gravity waves', *J. Atmos. Sci.*, **44**, pp. 1404-1410, 1987.

Smith, A. K., and L. V. Lyjak, 'An observational estimate of gravity wave drag from the momentum balance in the middle atmosphere', *J. Geophys. Res.*, **90**, pp. 2233-2241, 1985.

VanZandt, T. E., 'A universal spectrum of buoyancy waves in the atmosphere', *Geophys. Res. Lett.*, **9**, pp. 575-578, 1982.

Warner, C. D., and M. E. McIntyre, 'On the propagation and dissipation of gravity wave spectra through a realistic middle atmosphere'. *J. Atmos. Sci.*, **53**, pp. 3213-3235, 1996.

Warner, C. D., and M. E. McIntyre, 'An ultra-simple spectral parameterization for non-orographic gravity waves', *J. Atmos. Sci.*, **58**, pp. 1837-1857, 2001.

Yang, F., M. E. Schlesinger, E. V. Rozanov, N. Andronova, V. A. Zubov, and L. B. Callis, 'Sensitivity of middle atmospheric temperature and circulation in the UIUC GCM to the treatment of subgrid-scale gravity-wave breaking', *Atmos. Chem. Phys. Discuss*., **6**, pp. 9085-9121, 2006.