An Accurate Spectral Nonorographic Gravity Wave Drag Parameterization for General Circulation Models

JOHN F. SCI NOCCA

Canadian Centre for Climate Modelling and Analysis, Meteorological Service of Canada, Victoria, British Columbia, Canada

(Manuscript received 19 April 2002, in final form 2 September 2002)

ABSTRACT

An accurate method of solution is developed for the Warner and McIntyre parameterization of nonorographic gravity wave drag for the case of hydrostatic dynamics in the absence of rotation. The new scheme is sufficiently fast that it is suitable for operational use in a general circulation model. Multiyear climate runs of the Canadian Middle Atmosphere Model are performed using the new scheme, and these are compared against equivalent climate runs using the full Warner and McIntyre scheme, which includes nonhydrostatic and rotational wave dynamics. The results indicate that the new operational scheme closely reproduces the seasonal distribution of time- and zonal-mean zonal winds, and the seasonal evolution of lower stratospheric temperatures that are obtained when the full Warner and McIntyre scheme is used. In addition, these quantities are shown to compare favorably with observations.

1. Introduction

The zonal-mean wind and temperature structure of the middle atmosphere arises largely from a balance between radiative driving and gravity wave drag (GWD; e.g., Holton 1983). This balance requires gravity waves whose sources are both orographic and nonorographic. The term nonorographic refers to the fact that the sources of these waves are nonstationary and so induce waves with non-zero horizontal phase speeds. The drag exerted by nonorographic gravity waves is most important in the upper stratosphere and mesosphere. Their parameterization is therefore essential to general circulation models (GCMs) that include these regions.

In general, current parameterizations of nonorographic GWD do not account for the sources of these waves since the sources themselves require parameterization in GCMs (e.g., deep convection, boundary layer turbulence). The most common approach is one in which the source of nonorographic gravity waves is specified by imposing a “launch spectrum” in the troposphere or lower stratosphere that is typically independent of time and geographic location. This simplifies the problem so that one need only parameterize the vertical propagation of the wave field and the drag exerted by its nonlinear breakdown.

Recently, Warner and McIntyre (1996; hereafter WM96) have suggested a simple, albeit expensive, framework to parameterize nonorographic gravity wave drag that includes nonhydrostatic and rotational wave dynamics. The framework is simple in the sense that aspects of the parameterization problem that are clearly posed, such as the linear theory governing the conservative propagation of a field of gravity waves, are carefully distinguished from aspects that are more empirical in nature, such as the deposition of momentum arising from the nonlinear breakdown and turbulent dissipation of the wave field. Their approach is expensive primarily because it includes nonhydrostatic and rotational wave dynamics, which increase the complexity of the problem.

Currently, it is too costly to include nonhydrostatic and rotational wave dynamics in an operational parameterization of nonorographic GWD. Even so, it is important to have some understanding of their effects when interpreting results from operational schemes where they are neglected. To address this issue, Scinocca (2002; hereafter S02) optimized the method of solution for the WM96 scheme so that it could be used for a limited number of fully interactive multiyear climate simulations. These experiments indicated a weak dependence on rotation and a strong dependence on nonhydrostatic wave dynamics, which is summarized below.

For hydrostatic dynamics, the amount of momentum flux available to the flow aloft is identical to the amount...
of momentum flux prescribed in the imposed launch spectrum. For nonhydrostatic wave dynamics, on the other hand, the amount of momentum flux available to the flow aloft is generally less than the amount of momentum flux prescribed in the imposed launch spectrum. This is because nonhydrostatic wave dynamics allow portions of the upwardly propagating wave spectrum to back-reflect. As a consequence, one can define an “effective” launch spectrum, which is formed by the removal of those portions of the imposed launch spectrum that have experienced back-reflection. The effective launch spectrum can be very different in structure and significantly reduced in magnitude compared to the imposed launch spectrum. In S02 it was found that the effective launch spectrum displays a systematic seasonal and latitudinal variation that is due to the seasonal and latitudinal variation of the winds and temperatures in the middle atmosphere.

This seasonal variation of the launch spectrum is wholly absent in operational schemes that are hydrostatic. Even so, S02 demonstrated that nearly equivalent seasonal mean winds and temperatures can be obtained with a hydrostatic version of the WM96 scheme if one simply reduces the amount of launch momentum flux specified in each horizontal azimuth. This remarkable similarity suggests that there is great advantage to pursuing an operational version of the hydrostatic scheme presented in S02.

Such an operational scheme would require further simplification beyond the hydrostatic version of WM96 presented in S02 since that version was essentially as expensive as the full nonhydrostatic version. An obvious question that arises is, can an efficient operational scheme be derived that is formally identical to the hydrostatic scheme of S02? Here it is shown that the answer to this question is yes. Beginning with the full WM96 scheme we derive an efficient operational scheme that assumes hydrostatic wave dynamics in the absence of rotation. The new scheme is formally identical to, and explicitly validated against, the hydrostatic version of the WM96 scheme presented in S02. The new scheme is “operational” in the sense that it requires 4 to 8 times less computational time.

An operational version of the WM96 scheme, which also assumes hydrostatic dynamics in the absence of rotation, has been proposed by Warner and McIntyre (1999, 2001, hereafter WM99 and WM01 respectively). They have referred to this operational version as the “ultrasimple” scheme. In the ultrasimple scheme the dimensionality of the problem is reduced by integrating the launch spectrum with respect to the intrinsic frequency, rendering it a function of vertical wavenumber alone. The same approach is taken here but with some subtle but important differences. In the ultrasimple scheme the vertical wavenumber dependence of the spectrum on all levels is approximated by either two-part (WM99) or three-part (WM01) piecewise analytic functions. The operational scheme proposed here is derived in a way that avoids the need for such an approximation to the spectrum.

The outline of the paper is as follows: In section 2 the new scheme and its method of solution are presented. In section 3 the new scheme is validated against the hydrostatic version of WM96 presented in S02. In section 4 modifications to the launch spectrum designed to compensate for the absence of nonhydrostatic dynamics are investigated in a set of offline experiments. In section 5 a 5-yr present-day climate simulation using the new operational scheme is presented and compared to an equivalent simulation using the full nonhydrostatic version of the WM96 scheme. A brief summary is presented in section 6.

2. Operational hydrostatic scheme

a. Review of WM96 and S02

In this section we will briefly review the WM96 approach and the method of solution employed in S02 before outlining the new operational hydrostatic version of the scheme. The interested reader is directed to WM96 and S02 for a more complete description of the problem.

The wave parameters are the ground-based frequency \( \omega \) and the wavenumber \( k = (k_1, k_2, m) \), where \( k_1, k_2, \) and \( m \) are wavenumbers in the zonal, meridional, and vertical directions, respectively. Here we define \( k = (k_1^2 + k_2^2 + m^2)^{1/2} \) is the magnitude of the horizontal wavenumber and \( \phi \) is the azimuthal direction.

The well-known dispersion relation (Gill 1982) for incompressible flow that governs nonhydrostatic internal gravity waves in the presence of rotation may be written

\[
m^2 = \frac{k^2N^2 (1 - \hat{\omega}^2/N^2)}{\hat{\omega}^2 (1 - f^2/\hat{\omega}^2)},
\]

where \( N \) is the basic-state buoyancy frequency, \( f \) is the inertial frequency,

\[
\hat{\omega} = \omega - kU
\]

is the intrinsic frequency and \( U \) is the projection of the basic-state horizontal velocity onto the azimuthal direction \( \phi \). From (1) it can be seen that wavelike solutions \( (m^2 > 0) \) are defined only for intrinsic frequencies that fall in the range \( f^2 < \hat{\omega}^2 < N^2 \).

In this problem we assume an azimuthally isotropic launch spectrum which is independent of time and geographic location. In any azimuth \( \phi \), it is specified by the total wave energy per unit mass, the spectral density of which is assumed to be of the generalized Desaubies form (Fritts and VanZandt 1993):

\[
\hat{E}(m, \hat{\omega}, \phi) = C \left( \frac{m}{m_0} \right)^x \frac{N^2\hat{\omega}^{-p}}{1 - \left( \frac{m}{m_0} \right)^x},
\]
where \( m_\omega \) is a characteristic vertical wavenumber and \( C \) is a constant. This expression is separable in \( m \) and \( \omega \). The dependence on \( m \) is \( \dot{E} \propto m^{-t} \) for \( m \ll m_\omega \), and, \( \dot{E} \propto m^{s} \) for \( m \ll m_\omega \). Typical values for \( s \) and \( t \) are \( s = 1 \) and \( t = 3 \). The dependence on \( \omega \) is \( \dot{E} \propto \omega^{-p} \) for all \( \omega \). Observations (e.g., Fritts and VanZandt 1993) and physical arguments (e.g., WM96) suggest that values for \( p \) fall in the range \( 1 \leq p \leq 5/3 \).

The quantity of primary interest here is the spectral density of the vertical flux of pseudomomentum directed into the \( \phi \) azimuth. This quantity will be referred to as simply the Eliassen–Palm (E–P) flux density. The E–P flux density as a function of \( (m, \omega, \phi) \), referred to as \( m-\omega \) space, is obtained from \( \dot{E}(m, \omega, \phi) \) by the group velocity rule:

\[
\rho \dot{F}(m, \omega, \phi) = \rho \dot{c}_m \frac{k}{\omega} \dot{E}(m, \omega, \phi),
\]

where \( \rho \) is the basic-state density and \( \dot{c}_m = \partial \omega / \partial m \) is the vertical group velocity. The E–P flux density as a function of \( (k, \omega, \phi) \), referred to as \( k-\omega \) space, is obtained by application of the Jacobian of the transformation between the \( m-\omega \) and the \( k-\omega \) spaces:

\[
\rho F(k, \omega, \phi) = J_1 \rho \dot{F}(m, \omega, \phi),
\]

where the Jacobian, \( J_1 \), is given by

\[
J_1 = \frac{\partial (m, \omega, \phi)}{\partial (k, \omega, \phi)} = \frac{m}{k}.
\]

The launch spectrum, as represented by (3), (4), or (5), propagates vertically through height-varying basic-state wind \( U(z) \) and buoyancy frequency \( N(z) \). In this problem, the wind and buoyancy frequency are taken to be time-independent and horizontally uniform. As a result, the ground-based frequency \( \omega \) and horizontal wavenumber \( k \) are invariants. Therefore, spectral elements in \( k-\omega \) space \( (dkd\omega d\phi) \) are also invariant to changes in \( U \) and \( N \). On the other hand, the vertical wavenumber \( m \) and intrinsic frequency \( \omega \) are not invariant and change with changes in \( U \) and \( N \) through (1) and (2), respectively. Therefore, spectral elements in \( m-\omega \) space \( (dm\omega d\phi) \) are not invariant to changes in \( U \) and \( N \). Consequently, the density \( \rho F(k, \omega, \phi) \) is conserved for conservative propagation whereas the densities \( \rho \dot{F}(m, \omega, \phi) \) and \( \dot{E}(m, \omega, \phi) \) are not. This suggests that the \( k-\omega \) space has the most favorable conservation properties and is the natural coordinate frame in which to solve this problem, even though the dynamics are most straightforward in \( m-\omega \) space.

Schematics representing the appearance of the launch densities \( \rho \dot{F}(m, \omega, \phi) \) in \( m-\omega \) space and \( \rho F(k, \omega, \phi) \) in \( k-\omega \) space corresponding to (3) are presented in Fig. 1. The subscript \( o \) refers to the values of variables at the launch level. The solid contours in both frames represent the E–P flux densities while the shaded region indicates values of \( (m, \omega) \) and \( (k, \omega) \) for which wavelike solutions exist [i.e., \( m^2 > 0 \) in (1)]. The short-dashed contours in \( k-\omega \) space represent lines of constant vertical wavenumber \( m \), increasing in the direction indicated on the figure.

In \( m-\omega \) space, wavelike solutions exist only when the value of \( \omega \) falls within the bounds \( f \) and \( N \). In \( k-\omega \) space, from (2), these bounds on \( \omega \) are represented by the two lines \( \omega_c = f + kU \) and \( \omega_N = N + kU \). The line \( \omega_c \) corresponds to \( m \rightarrow \infty \) and so identifies wavenumber–frequency pairs that have their critical level at the current height level. The line \( \omega_N \) corresponds to \( m \rightarrow 0 \) and so identifies wavenumber–frequency pairs that undergo back-reflection at the current height level.

In the algorithm proposed by S02, the \( \omega_N \) and \( \omega_c \) lines are used to evaluate back-reflection and critical-level filtering respectively. As \( U \) and \( N \) change in the vertical, the two parallel lines \( \omega_c \) and \( \omega_N \) pivot back and forth about an anchor point on the \( k = 0 \) axis, removing
of the wave energy density (3) at each level. The saturated energy density spectrum used is
\[ \hat{E}_S(m, \tilde{\omega}, \phi) = C \left( \frac{m}{m_s} \right)^{-3} N^2 \tilde{\omega}^{-p} \]  
with \( 1 \leq p \leq 5/3 \). As discussed by WM96, the \( m^{-3} \) dependence has observational and theoretical support, while the \( \tilde{\omega}^{-p} \) dependence is less certain. From (3) and (7) we can see that, at launch, \( \tilde{E}(m, \tilde{\omega}, \phi) \) and \( \hat{E}_S(m, \tilde{\omega}, \phi) \) are equal at asymptotically large \( m \) when \( t = 3 \).

Expressions for E–P flux density in \( m-\omega \) space and \( k-\omega \) space corresponding to \( \tilde{E}_S(m, \tilde{\omega}, \phi) \) are obtained following (4) and (5). That is,
\[ \rho F_c(m, \tilde{\omega}, \phi) = \rho c_s \tilde{E}_S(m, \tilde{\omega}, \phi), \quad \text{and} \]
\[ \rho F_c(k, \omega, \phi) = J \rho F_c(m, \tilde{\omega}, \phi). \]

Therefore, all spectral elements that undergo conservative propagation are constrained by the saturation condition
\[ \rho F(k, \omega, \phi) \leq \rho F_c(k, \omega, \phi). \]

Further details and the full expressions for (4), (5), (8), and (9) may be found in S02.

b. The new operational scheme

The derivation of the new operational scheme from the full WM96 scheme is performed in three steps. In the first step we assume hydrostatic dynamics in the absence of rotation and obtain a new expression for the launch E–P flux \( \rho F^{H}(m, \tilde{\omega}, \phi) \). In the second step \( F^{H}(m, \tilde{\omega}, \phi) \) is transformed from \( m-\omega \) space to \( c-\tilde{\omega} \) space, where \( c = \omega k \) is the ground-based phase speed in the direction \( \phi \). In the third and final step, the \( \tilde{\omega} \) dependence of the E–P flux density is integrated out so that the density is a function only of the conserved variable \( c \) in each azimuth. There are significant differences between this approach and the approach taken to derive the ultrasimple scheme of WM99 and WM01. These differences will be discussed at the end of this section.

In the first step of the derivation we assume hydrostatic wave dynamics in the absence of rotation. This is achieved by taking the two limits
\[ \tilde{\omega}^2 N^2 \rightarrow 0 \quad \text{(hydrostatic wave dynamics)} \]
\[ f^2 \tilde{\omega}^2 \rightarrow 0 \quad \text{(no rotation)} \]
in all expressions. For example, using (11) and (12) in (1) we recover the usual dispersion relation for hydrostatic wave dynamics in the absence of rotation
\[ m^2 = \frac{k^2 N^2}{\tilde{\omega}^2} \left( \frac{N^2}{(c - U)^2} \right). \]

In addition, the expression (4) for \( \rho F^{H}(m, \tilde{\omega}, \phi) \) simplifies to

\[ \rho F_c(m, \tilde{\omega}, \phi) = \rho c_s \tilde{E}_S(m, \tilde{\omega}, \phi), \quad \text{and} \]
\[ \rho F_c(k, \omega, \phi) = J \rho F_c(m, \tilde{\omega}, \phi). \]
\[ \rho \hat{F}(m, \omega, \phi) = \frac{k}{m} \hat{E}(m, \omega, \phi), \]

(14)

since \( c' = \omega/m \).

Following S02, we shall refer to nonhydrostatic wave dynamics in the presence of rotation as the N + R dynamics system, and hydrostatic dynamics in the absence of rotation as the H − R dynamics system. The two systems are illustrated in Fig. 2 where \( k−\omega \) space schematics of the E−P flux density at launch (solid contours) are illustrated for N + R dynamics at the top and H − R dynamics at the bottom. Dashed contours in both correspond to contours of constant \( m \). There are differences between the two systems and their method of solution; these are discussed at length in S02.

It is important to note, that for the N + R system, the E−P flux density is restricted to intrinsic frequencies in the range \( f^2 < \omega^2 < N^2 \). For the H − R system, the E−P flux density may be nonzero for intrinsic frequencies in the range \( 0 < \omega^2 < \infty \). In order to make meaningful comparisons between the two systems, as indicated in Fig. 2, the launch spectrum in both systems will be restricted to intrinsic frequencies in the range \( f^2 < \omega^2 < N^2 \).

In the second step of the derivation of the new operational scheme, the E−P flux density in \( c-\omega \) space is derived from \( \hat{F}(m, \omega, \phi) \) via

\[ \rho F(c, \omega, \phi) = J_2 \rho \hat{F}(m, \omega, \phi), \]

(15)

where the Jacobian, \( J_2 \), is given by

\[
\rho F(c, \omega, \phi)|_{z = z_o} = \begin{cases} 
\rho C \left( \frac{m}{m_o} \right)^{\frac{1}{2}} \frac{m^2}{1 + \left( \frac{m}{m_o} \right)^{\frac{2}{s}}}, & \omega_{lw}(z_o) \leq \omega \leq \omega_{hi}(z) \\
0, & \text{otherwise}
\end{cases}
\]

(18)

In (18) we have used \( \omega_{lw}(z) \) and \( \omega_{hi}(z) \) to represent the Doppler-shifted bounds on the region of nonzero \( F(c, \omega, \phi) \). On the launch level, \( z_o \), we have \( \omega_{lw}(z_o) = f \) and \( \omega_{hi}(z_o) = N \). Formally, the integration of (17) with respect to \( \omega \) must occur over the range \( 0 \leq \omega \leq \infty \).

From (18), on the launch level, this is equivalent to integrating (17) over the range \( f \leq \omega \leq \infty \). On subsequent levels, \( z > z_o \), this is equivalent to integrating (17) over the range \( \omega_{lw}(z) \leq \omega \leq \omega_{hi}(z) \). Integrating (17) with respect to \( \omega \) we may write

\[ \rho \bar{F}(c, \phi) = \rho C \bar{I}(z; \phi) \left( \frac{m}{m_o} \right)^{\frac{1}{2}} \frac{m^2}{1 + \left( \frac{m}{m_o} \right)^{\frac{2}{s}}}, \]

(19)

where the overbar denotes integration with respect to \( \omega \) and

\[ I(z; \phi) = \int_{\omega_{lw}(z)}^{\omega_{hi}(z)} \tilde{\omega}^{1-p} d\tilde{\omega} 
= \frac{\omega_{hi}(z)^{2-p} - \omega_{lw}(z)^{2-p}}{2 - p}. \]

(20)

It remains, then, to determine the values of \( \omega_{lw}(z) \) and \( \omega_{hi}(z) \) given basic-state profiles of \( U(z) \) and \( N(z) \). For a given \( \phi \), \( \omega_{lw}(z) = k(c - U(z)) \) denotes the intrinsic frequency on the launch level \( z_o \) of a spectral element with horizontal wavenumber \( k \) and ground-based frequency \( \omega \) (or phase speed \( c = \omega/k \)). The same spectral element at a higher elevation will have the intrinsic
frequency $\omega(z) = k(c - U(z))$. The ratio $\dot{\omega}(z)/\omega(z)$ implies the relation:

$$
\dot{\omega}(z) = \frac{\omega(z)}{c - U(z)} \left[ \frac{c - U(z)}{c - U(z_0)} \right].
$$

(21)

Using (21) to relate $\dot{\omega}_{\text{low}}(z)$ and $\dot{\omega}_{\text{hi}}(z)$ to their respective values of $\dot{\omega}_{\text{low}}(z_0) = f$ and $\dot{\omega}_{\text{hi}}(z_0) = N_o$ on the launch level we may write (20) as

$$
I(z; p) = \left( \frac{c - U}{c - U_o} \right)^{2-p} \left[ \frac{N_o^{2-p} - f^{2-p}}{2 - p} \right].
$$

(22)

The E–P flux density for the present operational scheme is then given by (19) and (22). One of the most important input parameters of the new scheme is the total integrated E–P flux directed into each azimuth at launch. Following S02 this is referred to as $\rho F^\text{total}_p$. The specification of $\rho F^\text{total}_p$ sets the value of $C I(z; p)$ in (19) at the launch level. The importance of respecting the formal limits of 0 and $\infty$ in the integration of (17) with respect to $\omega$ is actually only realized in the normalization of the saturated spectrum [(26) below]. Since the launch spectrum and saturated spectrum are equivalent at asymptotically large $m$, the normalization coefficient of the saturated density varies in the vertical like

$$
\frac{I(z; p)}{I(z_o; p)} = \left( \frac{c - U}{c - U_o} \right)^{2-p}.
$$

(23)

This vertical variation is independent of the spectrum on any level and is simply a consequence of the Doppler-shifting of $\omega$ with height. Had we not taken the Doppler-shifting of $\omega$ in account and simply employed $\dot{\omega}_{\text{low}}(z) = f$ and $\dot{\omega}_{\text{hi}}(z) = N_o$ on all levels, then the right-hand side of (23) would be unity. The correct application of the saturation condition, therefore, depends critically on accounting for the Doppler-shifting of the limits of integration $\dot{\omega}_{\text{low}}(z)$ and $\dot{\omega}_{\text{hi}}(z)$ in (20).

It is convenient to perform the Galilean transformation:

$$
\tilde{U} = U - U_o, \quad \tilde{c} = c - U_o
$$

(24)

and express the E–P flux density (19) as:

$$
\rho F(\tilde{c}, \phi) = \rho \tilde{A} \frac{\tilde{c} - \tilde{U}}{N} \left( \frac{\tilde{c} - \tilde{U}}{\tilde{c}} \right)^{2-p} \frac{1}{1 + \left( m(\tilde{c} - \tilde{U})/N \right)^{s+3}},
$$

(25)

where

$$
\tilde{A} = C m \left[ \frac{N^{2-p} - f^{2-p}}{2 - p} \right].
$$

The transformation (24) renders the launch E–P flux independent of azimuth since we have $\tilde{U} = 0$ on this level. In (25) we have used $t = 3$, eliminated $m$ with the dispersion relation (13), and grouped together all terms that are independent of height into a new constant $A$.

The advantage gained by solving the $H - R$ system in terms of the conserved wave variable $\tilde{c}$ (or equivalently $c$) is analogous to the advantage gained by solving the $N + R$ system in terms of the conserved wave variables $k$ and $\omega$; since these wave variables are conserved, the E–P flux density is conserved for conservative propagation. Consequently, the density is altered only through dissipative processes associated with critical-level filtering and saturation.

In Fig. 3 schematics of the launch E–P flux density in $m$ space and in $\tilde{c}$ space are illustrated for a given azimuth. In this figure we have used $t = 3$ and presented curves for the two cases $s = 1$ and $s = 0$. The two E–P flux densities $\rho F(\tilde{c}, \phi)$ and $\tilde{\rho} F(\tilde{c}, \phi)$ are related to each other by $J_2$ since (16) is not altered as a consequence of the $\omega$ integration. It can be seen that the two densities have different structure in each space. As the spectrum propagates conservatively through varying winds and buoyancy frequency, the $m$-space representation of the density deforms while the $\tilde{c}$-space representation of the density remains unchanged.

The application of critical-level filtering is straightforward. In each azimuth the wind $U = U_o$ at launch implies $\tilde{U}(z_0) = 0$. This sets the absolute lower bound
of $\tilde{c} = 0$ for nonzero values of $\tilde{F}(\tilde{c}, \phi)$. If, on the next level above the launch level, $\tilde{U}$ increases to a value of $\tilde{U}(z_t) > \tilde{U}(z_t)$, then waves with phase speeds in the range $\tilde{U}(z_t) \leq \tilde{c} \leq \tilde{U}(z_t)$ encounter critical levels between $z_t$ and $z_t$. The momentum flux between these phase speeds is removed from the density $\tilde{F}(\tilde{c}, \phi)$ and this amount is deposited to the flow in this layer. The same procedure is applied on all subsequent levels and in all azimuths.

The application of saturation follows S02. Following the derivation of (25) we may derive a saturated spectrum from (8) of the form

$$\tilde{F}^s(\tilde{c}, \phi) = \rho \tilde{A} \left( \tilde{c} - \tilde{U} / \tilde{c} \right)^{2-p}. \quad (26)$$

For each spectral element that experiences conservative propagation from one level to the next, the saturation condition

$$\tilde{F}^s(\tilde{c}, \phi) \leq \tilde{F}^s(\tilde{c}, \phi) \quad (27)$$

is used to limit the value of $\tilde{F}^s(\tilde{c}, \phi)$. The total integrated momentum flux removed by the application of (27) in each model layer represents the amount of momentum deposited to the flow in the current azimuth due to saturation.

The method of solution for the new scheme is as follows: The azimuthal dependence of the launch spectrum is represented by a discrete number $n_c$ of equally spaced azimuths $\phi$. In each azimuth, the launch spectrum $\tilde{F}^s(\tilde{c}, \phi)$ is discretized in horizontal phase speed $\tilde{c}$ by a fixed number of spectral elements $n_c$. As described earlier, the integrated $E-P$ flux directed into each azimuth at launch, $\tilde{F}^s_{total}$, is a free parameter in the problem and it is used to normalize $\tilde{F}^s(\tilde{c}, \phi)$. In the propagation of the spectrum from one level to the next, dissipative processes are represented by the application of critical-level filtering and by the application of the saturation bound (27). The ultimate product of this effort is two profiles of net eastward and northward vertical E-P flux. The vertical divergence of these fluxes provide the zonal and meridional wind acceleration induced by the dissipation of nonorographic gravity waves that make up the launch spectrum.

In order to optimize the performance of the new scheme, following S02, a coordinate stretch has been applied to $\tilde{c}$ in each of the $n_c$ azimuthal directions. The coordinate stretch is used to increase resolution of the E-P flux density at large phase speeds (i.e., small vertical wavenumbers), since this portion of the spectrum is most crucial to the drag in the mesosphere. The problem is solved in the space of the transformed variable $X$ which has uniform resolution $dX$. Here we take the untransformed variable $X$ to be

$$X = \frac{1}{\tilde{c}}, \quad (28)$$

which is proportional to the launch vertical wavenumber. The coordinate stretch is defined by

$$X = B_1 e^{X - X_{max}/\Gamma} + B_2 \quad (29)$$

$$\frac{dX}{d\tilde{X}} = B_1 \Gamma e^{(X - X_{max})/\Gamma}, \quad (30)$$

where $X_{min} \leq X \leq X_{max}$. Imposing the constraint $X_{min} \leq X \leq X_{max}$ implies

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

$$B_2 = X_{min} - B_1.$$

The free parameters in this transformation are $X_{min}$ and $X_{max}$, which set the values of $X_{min}$ and $X_{max}$, respectively, and the half-width of the stretch $\Gamma$. Since the above transformation is independent of azimuth and geographic location, it need only be calculated once. A more detailed description of the transformation (29) can be found in S02.

There are several important differences between the operational scheme derived here and the ultrasilpme operational scheme developed by WM99 and WM01. While WM99 and WM01 also integrate out the $\phi$ dependence of the E-P flux density to render it one-dimensional in each azimuth, they choose to express it as a function of vertical wavenumber $m$ rather than phase speed $c$ as is done here. Consequently, even though total E-P flux is conserved, their representation of the E-P flux density is not since $m$ is not conserved for conservative propagation. To deal with this, WM99 approximates the E-P flux density on all levels by a two-part function. The first part of the function represents the portion of the E-P flux density that has conservatively propagated from launch and it may be integrated analytically by mapping this portion back down to the launch level. The second part of the function represents the portion of the spectrum that has experienced saturation and it may be integrated analytically by mapping this portion of the spectrum back down to the level where saturation was imposed. Attempts to alleviate some of the artifacts associated with approximating the spectrum by a two-part function were undertaken by WM01, where this procedure was generalized to an approximation consisting of a three-part function.

Another important difference between the present approach and that of WM99 and WM01 is the way in which the integration of $\phi$ is performed. Here, the E-P flux density is first rendered hydrostatic and rotation eliminated (i.e., H-R dynamics) before the integration over $\phi$ was performed. In WM99, in order to facilitate offline comparisons of their ultrasilpme two-part scheme with the full WM96 scheme, the integration over $\phi$ was performed numerically on the full $N + R$ form of the E-P flux density from WM96. This integral was then used to normalize the E-P flux density of their H-R two-part scheme.
In order to use the ultrasimple scheme as a parameterization in a GCM, WM01 proposed an estimate of the integral of the E–P flux density over \( \omega \) (i.e., their appendix D). This estimate amounted to rendering the E–P flux density hydrostatic and removing rotation prior to the integration over \( \omega \). This is more like the approach taken here. However, in evaluating this integral for H–R dynamics, WM01 employ the integration limits of \( \omega_{\text{min}}(z) = f \) and \( \omega_{\text{max}}(z) = N \) on all levels. As discussed earlier in this section, this will result in an error in the application of the saturation condition in their scheme. This error is investigated further in the next section.

### 3. Validation of the new scheme

In this section we validate the new scheme against the H–R version of WM96 presented in S02. Since the E–P flux density \( \times \omega^{1/2} \) for H–R dynamics [see (17)], it becomes independent of \( \omega \) when \( p = 1 \). In S02 this was taken to imply that an operational scheme that eliminated the \( \omega \) dependence of the E–P flux density through integration would be formally identical to the H–R version of WM96 when \( p = 1 \). This result is actually more general and it requires clarification.

The spectral density of the launch spectrum and the spectral density of the saturated spectrum selected for the S02 study and the present study have the property that, in the H–R limit, their dependence on \( \omega \) is identical (i.e., the E–P flux density \( \times \omega^{1/2} \) for both). This means that the application of saturation is independent of \( \omega \). In any azimuth it depends on \( m \) and on the vertical variation of the normalization coefficient (23). Further, since the launch spectral density is assumed to be separable in \( \psi \) and \( m \), critical-level filtering, which depends only on the Doppler shifting of \( m \to \infty \), also occurs independently of \( \omega \). Therefore, in the H–R limit, one is free to first integrate the \( \omega \) dependence out of the launch and saturated E–P flux densities to reduce the dimensionality of the problem. Consequently, the H–R system of S02 and the operational scheme developed here are formally identical for all \( p \). An explicit comparison of the two schemes, therefore, provides an excellent test to validate the new scheme.

Here, we compare the S02 H–R scheme against the new operational scheme for the two values of \( p = 1 \) and \( p = 1.5 \), which were used extensively by S02. For this comparison the values of the free parameters in each scheme are taken to be: \( s = 1 \), \( t = 3 \), \( n_\psi = 4 \) (with cardinal directions of N, S, E, W), \( m_\psi = 2 \) m/2000 m\(^{-1} \), \( \rho F_\psi^{\text{total}} = 3.54 \times 10^{-4} \) Pa and a launch elevation of 125 hPa. Additional parameters required for the S02 implementation of the WM96 scheme are: \( n_k = 512 \), \( n_\omega = 512 \), \( \Gamma_k = 1.4 \) m\(^{-1} \), \( \Gamma_\omega = 0.2(\omega^k - \omega^c) \), and \( k_{\text{max}} = 1.0 \times 10^{-1} \) m\(^{-1} \) (further details regarding these quantities may be found in S02). Additional parameters required for the new operational scheme are: \( n_k = 1000 \), \( c_{\text{min}} = 0.25 \) m s\(^{-1} \), \( c_{\text{max}} = 2000 \) m s\(^{-1} \), and \( \Gamma = 0.6 \) s m\(^{-1} \). It is important to note that the E–P flux density is expressed as a function of \( k \) and \( \omega \) in the S02 implementation and by \( c = \omega k \) in the new operational scheme. Because each scheme spans wavenumber–frequency space differently, the two schemes are formally identical only when \( 0 \leq k \leq \infty \) and \( 0 \leq c \leq \infty \). Consequently, for this comparison the value of \( k_{\text{max}} \) has been taken to be an order of magnitude larger than the value used in S02.

Midlatitude profiles of \( U \) and \( N \), which are representative of summer and winter seasons, are extracted from a previous GCM simulation and used as input for the two schemes in an offline mode. These profiles are plotted in Fig. 4. The horizontal dashed line in this figure indicates the launch elevation used in each scheme. Profiles of horizontal wind acceleration for each scheme are then compared for consistency. Profiles of zonal- and meridional-wind accelerations (or drag), \( U_i \) and \( V_i \), resulting from this set of experiments are plotted in Fig. 5 above 20-km elevation. The thick gray line corresponds to the S02 H–R scheme, while the thin black line corresponds to the new operational scheme. The two schemes produce nearly identical results for both \( p = 1 \) and \( p = 1.5 \).

In order to quantify the effect of employing the correct integration bounds in (20), additional acceleration profiles (dashed lines) have been added to Fig. 5 that are derived using the integration bounds \( \omega_{\text{min}}(z) = f \) and \( \omega_{\text{max}}(z) = N \). As discussed earlier, this amounts to eliminating the vertical variation of the saturated spectrum (23), which arises from Doppler shifting. The zonal drag is most strongly affected by this change. In winter, the deposition of easterly momentum begins at lower elevation. Due to this premature depletion, there is less easterly momentum available in the wave field at greater elevation. Consequently, the peak of easterly acceleration of the mean flow near 55 km is underestimated by roughly a factor of 2 for \( p = 1.5 \) and nearly a factor of 5 for \( p = 1 \). In summer, the deposition of westerly momentum also begins at an elevation that is too low. For similar reasons, the peak of westerly acceleration of the mean flow near 75 km is also underestimated by roughly a factor of 2 for \( p = 1.5 \) and a factor of 8 for \( p = 1 \). It is clear from Fig. 5, then, that failure to account for Doppler shifting by the basic flow when integrating the saturated spectrum with respect to \( \omega \) can result in a substantial error in the profile of acceleration produced by the parameterization scheme.

### 4. Adjustment of launch spectrum

In S02 fully interactive 5-yr climate runs of the WM96 scheme employing N + R and H–R dynamics were undertaken to investigate the effect of including nonhydrostatic and rotational dynamics. In each of these runs everything was held constant except for \( \rho F_\psi^{\text{total}} \).
which was reduced by a factor of 4 in the H – R run to crudely reproduce the reduction caused by back-reflection in the N + R run. The time-mean zonal-mean zonal wind for July from the two runs is reproduced here in Fig. 6. A comparison of the two indicates a remarkable similarity of the climatological winds between the N + R and H – R runs. The primary difference between the two runs appears to be related to an unrealistically weak wind reversal at the summer mesopause in the H – R run. The corresponding comparison for January (not shown) indicates a similar bias (see S02).

As demonstrated in the previous section, the new operational scheme and the H – R scheme of S02 produce identical results. Consequently, the new operational scheme would reproduce the bias identified in Fig. 6. In this section, we consider modifications to the launch spectrum for H – R dynamics that are designed to alleviate this bias. Rather than making the launch spectrum more realistic, these modifications should be viewed simply as compensating for the neglect of back-reflection in the H – R system.

Differences in the N + R and H – R versions of WM96, which produced the winds in Fig. 6, are nicely highlighted when each is applied to identical wind and buoyancy frequency profiles in an offline mode. The zonal wave drag produced by the application of these two versions of WM96 on the representative summer and winter profiles of Fig. 4 are displayed in Fig. 7. In summer it is clear that the N + R system (solid line) produces more drag at the mesopause than the H – R system (long-dashed line). This is consistent with the unrealistically weak mesopause wind reversal obtained for the H – R run in Fig. 7. In winter, the zonal drag produced with H – R dynamics is everywhere smaller in magnitude compared to that produced with N + R dynamics. This is primarily due to the fact that back-reflection is not as important for winter conditions (see S02) and the weaker drag is a direct result of the factor of 4 reduction in \( \rho F_p \) for H – R dynamics.

A number of offline experiments were performed in which parameters governing the properties of the launch spectrum were adjusted in the H – R system to bring its drag profiles in Fig. 7 more in line with the N + R drag profiles. The parameters adjusted in this set of experiments are: \( \rho F_p^{\text{total}} \), the exponent \( s \) that governs the low-\( m \) portion of the launch spectrum, and \( m \) the characteristic vertical wavenumber where the launch spectrum (3) changes from an \( m^s \) dependence (\( m \ll m \)) to an \( m^{-1} \) dependence (\( m \gg m \)). For reference, the values of these parameters used in the H – R run of S02 are \( \rho F_p^{\text{total}} = 7.0 \times 10^{-4} \) Pa, \( s = 1 \), and \( m = 2\pi/2000 \) m\(^{-1} \). After a number of experiments, an optimal set of values for these parameters was found to be: \( \rho F_p^{\text{total}} = 4.25 \times 10^{-4} \) Pa, \( s = 0 \), and \( m = 2\pi/1000 \) m\(^{-1} \). The H – R zonal drag resulting from this new launch spectrum is given by the thick gray line in Fig. 7. The new operational scheme was used for this application with \( n_x = 1000 \) spectral elements of horizontal phase speeds. We can see that the modifications to the launch spectrum have now increased the drag in the summer mesosphere so that it roughly equals the N + R drag (solid line)
Fig. 5. Profiles of horizontal wind acceleration resulting from the application of the H – R scheme of S02 (thick gray lines) and the new operational scheme (thin black lines) on the wind and buoyancy frequency profiles in Fig. 4. Each scheme used identical parameter settings and results are presented for both \( p = 1 \) and \( p = 1.5 \). Dashed lines represent the wind accelerations obtained when Doppler shifting is not included in the integral in (20).

above 80 km. The zonal drag in winter for the H – R modified launch spectrum case is also closer to the \( N + R \) curve above 55 km.

As in S02, in order to optimize the new scheme for use in a GCM, a number of self-consistency tests were performed where the number of spectral elements \( n_c \) used to resolve the launch spectrum in each azimuth was reduced and an optimal stretching parameter \( \Gamma \) in the transformation (30) determined. These sensitivity tests indicated that \( n_c = 35 \) with \( \Gamma = 0.6 \text{ m s}^{-1} \) is sufficient to resolve the E–P flux density. In Fig. 7 the drag profile for this case (short-dashed line) is displayed alongside the well-resolved \( n_c = 1000 \) case (thick gray line). A comparison of the two indicates that \( n_c = 35 \) with \( \Gamma = 0.6 \text{ m s}^{-1} \) is sufficient resolution. These values will be used in the next section where GCM experiments are undertaken with the new scheme.

5. Multiyear GCM simulations

The new operational scheme has been implemented in the latest version of the Canadian middle atmosphere model (CMAM; initially reported by Beagley et al. 1997). This model is based upon the Canadian Centre for Climate Modelling and Analysis (CCCma) third-generation atmospheric GCM (AGCM3), which is similar in many respects to the second-generation model described in McFarlane et al. (1992). The model is spectral and, for the experiments discussed here, it employs triangular truncation at total wavenumber of 47 (T47). In the vertical, the model domain extends up to approximately 100 km and is spanned by 59 layers. The layer depths increase monotonically with height through the troposphere from approximately 100 m at the surface to approximately 3 km in the lower stratosphere and remain constant at this value above.
New or improved features of the parameterized physical processes in AGCM3 are: a new land surface scheme (CLASS; Verseghy et al. 1993), a new parameterization of cumulus convection (Zhang and McFarlane 1995), an improved treatment of solar radiation which employs four bands in the visible and near infrared, an “optimal” spectral representation of the earth’s topography (Holzer 1996), a revised representation of turbulent transfer coefficients at the surface (Abdella and McFarlane 1996), and a new anisotropic orographic GWD parameterization (Scinocca and McFarlane 2000).

The GCM experiments presented here are comprised of a series of multiyear simulations of present-day climate. These experiments are forced by an Atmospheric Model Intercomparison Project II (AMIPII) ensemble-mean climatology of monthly mean sea surface temperature and sea ice extent. That is, there is no interannual variability associated with this forcing. The model is spun up from observations and then integrated for 5 yr. Monthly mean quantities are derived from ensemble averages over the 5-yr simulations. These experiments will be referred to as 5-yr climate runs.

A summary of the parameters used in the new operational scheme for this set of experiments is as follows:

---

**Fig. 6.** Time-mean zonal-mean zonal winds (m s⁻¹) for Jul from the (upper) N + R and (lower) H – R 5-yr present-day climate simulations of S02. Progressively darker shading has been used to indicate regions of progressively increasing wind magnitude. (lower) The neglect of nonhydrostatic dynamics results in a wind bias in which the wind reversal at the summer mesopause is too weak.
FIG. 7. Profiles of zonal wind acceleration resulting from the application of the N + R (solid line) and H - R (long-dashed line) schemes of S02 and the new scheme using a modified launch spectrum with $n_c = 1000$ (thick gray lines) and $n_c = 35$ (short-dashed lines) spectral elements. The profiles of wind and buoyancy frequency in Fig. 4 were employed.

$p = 1.5$  \( \hat{\omega} \) exponent in (3) and (25)
$s = 0$  low-\( m \) exponent in (3) and (25)
$n_\phi = 4$  number of discrete azimuths (N, S, E, W)
$n_z = 35$  number of discrete intrinsic phase speeds in each \( \phi \)
$\Gamma = 0.6$ s m$^{-1}$  half-width of coordinate stretch
$m_* = (2\pi/1000)$m$^{-1}$  characteristic vertical wavenumber
$\bar{c}_{\text{min}} = 0.25$ m s$^{-1}$  minimum launch intrinsic phase speed in each \( \phi \)
$\bar{c}_{\text{max}} = 2000$ m s$^{-1}$  maximum launch intrinsic phase speed in each \( \phi \)
$\rho_p F_p^{\text{total}} = 4.25 \times 10^{-4}$ Pa  total launch E-P flux in each \( \phi \)
$P_{\text{launch}} = 125$ hPa  launch elevation.

In Fig. 8 we present the time-mean zonal-mean zonal winds for January and July from observations and from the two 5-yr climate runs. The observed winds, displayed in Figs. 8a,b, are derived from the Cooperative Institute for Research in the Atmosphere (CIRA) dataset (Fleming et al. 1990). Figs. 8c,d present the 5-yr averaged monthly mean winds from a run employing the full N + R WM96 version of the nonorographic GWD parameterization presented in S02. Figs. 8e,f present the 5-yr averaged monthly mean winds from a run using the new operational scheme. The N + R simulation compares favorably with the CIRA winds in both seasons. As discussed in S02, the use of the full WM96 scheme in the CMAM alleviates a number of wind biases that were present in the CMAM when the Hines (1997a,b) Doppler-spread nonorographic GWD parameterization was used.

A comparison of the winds in Fig. 8 resulting from the application of the new operational scheme (Figs. 8e,f) with the winds from the N + R run indicate a close correspondence. By modifying the launch spectrum for H - R dynamics in the new scheme, we have been able to alleviate much of the mesopause wind reversal bias present in Fig. 6.

There remain a number of wind biases in Fig. 8. For example, there is the usual bias that the SH winter jet does not tilt as much as observed. This bias is presumably related to the fact that current gravity wave drag parameterizations are still too crude. Fluctuations observed in the lower tropical stratosphere of the N + R run (Figs. 8c,d) are related to quasi-biennial oscillations (QBO)-like oscillations, which are driven as a result of the greater launch momentum flux in this run. This is discussed at length in S02. The westerlies in the lower stratosphere of the Northern Hemisphere during winter are too weak. This bias is also present on the tropospheric version of the model (e.g., see Scinocca and McFarlane 2000) and appears to be unrelated to the parameterization of nonorographic gravity waves.

The new operational scheme is substantially more efficient than the optimized N + R scheme developed in S02. Application of the N + R scheme in S02 was
found to double the required computational (cpu) time. The new operational scheme employing the previous parameters (31) increases the cpu time of the GCM by approximately 25%. This represents a factor of 4 savings. Subsequent tests indicate that this cost can be reduced to a 13% increase in cpu time with similar results (not shown). This additional savings may be achieved by reducing the number $n_c$ of horizontal phase speeds used to represent the E–P flux density in each azimuth from 35 to 15, reducing the maximum phase speed from 2000 m s$^{-1}$ to 100 m s$^{-1}$, and, reducing $\Gamma$ from 0.6 to 0.25 (m s$^{-1}$). For reference, the Hines Doppler-spread parameterization (Hines 1997a,b) in the CMAM increases the cpu time by 18%.3

In Fig. 9 we consider the seasonal evolution of the zonal-mean temperature in the lower stratosphere (50-hPa level) in these experiments relative to assimilated data. At the top of Fig. 9, this quantity is plotted for an ensemble of 8 yr (1994–2001) of U.K. Met. Office

---

3 This estimate refers to an application of the Hines scheme with constant launch E–P flux density at all $m$, and with no iteration performed on the cutoff wavenumber on each level. Altering either of these could significantly increase the cost.
(UKMO) operational assimilated data (Swinbank and O’Neill 1994). In Fig. 9a the middle and bottom, this quantity is plotted for the full N + R S02 simulation and for the run using the new operational scheme, respectively. The two simulations display a close similarity to each other and compare favorably to the UKMO assimilated data. The most obvious difference is a winter warm pole bias in the simulations.

A number of factors could account for this bias. The cold spring vortex seen in the UKMO data is aided by radiative cooling associated with the ozone loss due to heterogeneous chemical processes. These chemical pro-
cesses are not modeled in the simulations. Rather, radiative heating due to ozone is determined by a specified ozone climatology. Here we have employ the middle atmosphere dataset of Kita and Sumi (1986). This older climatology does not contain an ozone hole and, therefore, the associated cooling in the SH lower stratosphere during late spring is not represented in the simulations.

The warm pole bias throughout the winter appears in part due to a weakening of the polar vortex arising from the new orographic gravity wave drag parameterization scheme (Scinocca and McFarlane 2000), which produces more drag in the stratosphere than the orographic scheme used previously (McFarlane 1987). Sensitivity tests in which the strength of the orographic GWD is reduced (not shown) result in a stronger SH vortex and colder, more realistic, temperatures throughout winter. The warm bias evident in the NH lower stratosphere during winter is less sensitive to changes in the strength of the orographic gravity wave drag. It similarly occurs in a tropospheric version of the model (Scinocca and McFarlane 2000).

6. Summary

In this paper we have developed an operational version of the Warner and McIntyre (1996; WM96) spectral parameterization of nonorographic gravity wave drag for hydrostatic dynamics in the absence of rotation (H – R dynamics). Central to the derivation of the new operational scheme is the simplification of the H – R form of the E–P flux from \( \tilde{\mathbf{E}}(m, \phi) \) (used in S02) to \( \tilde{\mathbf{E}}(m, \phi) \) by integration over \( \phi \). This simplification reduces the required computational time by a factor of 4 to 8. Here, we have demonstrated that, when the Doppler shifting of \( \phi \) is correctly accounted for in this integration, the use of \( \tilde{\mathbf{E}}(m, \phi) \) is formally identical to the use of \( \tilde{\mathbf{E}}(m, \phi, \phi) \) in the parameterization problem.

This identity implies that all comparisons made between the N + R scheme and the H – R scheme in S02 are equally valid for the the new operational scheme as well. In S02 it was found that simply reducing the total amount of E–P flux launched with H – R wave dynamics could crudely mimic the effects of back-reflection obtained when nonhydrostatic and rotational (N + R) wave dynamics were used in the scheme. In this way, the H – R run of S02 was able to reproduce much of the seasonal-mean zonal-mean zonal wind structure of the N + R run. The main bias appeared to be an unrealistically weak wind reversal at the summer mesopause in the H – R run.

Here, using the new operational scheme, additional modifications to the launch spectrum were undertaken to make the H – R system better compensate for the absence of back-reflection. Five-year present-day climate simulations employing the new operational scheme revealed that these modifications to the launch spectrum were largely successful in eliminating this wind bias.

Therefore, one of the main results of this analysis is that the monthly mean zonal-mean winds and temperatures obtained from use of the full N + R WM96 scheme can be well reproduced by an H – R “operational” scheme.

In addition, it was found that the seasonal evolution of the zonal-mean temperature on the 50-hPa level was very similar between the full WM96 scheme (N + R) and the H – R operational scheme, and that both compared favorably with UKMO assimilated data. A modest winter warm pole bias was identified in the simulations. This bias is thought to be associated with too much orographic gravity wave drag and the absence of heterogeneous chemical processes in the simulations.

A portable version of the new operational scheme derived here, and the N + R WM96 scheme presented in S02 are freely available at http://www.cccma.bc.ec.gc.ca/~jscinocca/.

Acknowledgments. The author thanks Charles McLandress, John Fyfe, and an anonymous reviewer for their careful reading of the manuscript and helpful comments.

REFERENCES


