Mean-Flow Effects of Thermal Tides in the Mesosphere and Lower Thermosphere

ERICH BECKER
Leibniz Institute of Atmospheric Physics, Kühlungsborn, Germany

(Manuscript received 27 June 2016, in final form 24 March 2017)

ABSTRACT

This study addresses the heat budget of the mesosphere and lower thermosphere with regard to the energy deposition of upward-propagating waves. To this end, the energetics of gravity waves are recapitulated using an anelastic version of the primitive equations. This leads to an expression for the energy deposition of waves that is usually resolved in general circulation models. The energy deposition is shown to be mainly due to the frictional heating and, additionally, due to the negative buoyancy production of wave kinetic energy. The frictional heating includes contributions from horizontal and vertical momentum diffusion, as well as from ion drag. This formalism is applied to analyze results from a mechanistic middle-atmosphere general circulation model that includes energetically consistent parameterizations of diffusion, gravity waves, and ion drag. This paper estimates 1) the wave driving and energy deposition of thermal tides, 2) the model response to the excitation of thermal tides, and 3) the model response to the combined energy deposition by parameterized gravity waves and resolved waves. It is found that thermal tides give rise to a significant energy deposition in the lower thermosphere. The temperature response to thermal tides is positive. It maximizes at polar latitudes in the lower thermosphere as a result of poleward circulation branches that are driven by the predominantly westward Eliassen–Palm flux divergence of the tides. In addition, thermal tides give rise to a downward shift and reduction of the gravity wave drag in the upper mesosphere. Including the energy deposition in the model causes a substantial warming in the upper mesosphere and lower thermosphere.

1. Introduction

The dynamical control of the middle atmosphere is mainly due to waves that are generated in the troposphere and then propagate to higher altitudes where they become dynamically unstable, break, and cause wave–mean-flow interactions. Nonorographic gravity waves and thermal tides (herein “tides”) can propagate over many scale height densities before they break or are attenuated by other dissipative processes. Therefore, these waves can generate quite strong mean-flow effects per unit mass in the mesosphere and lower thermosphere (MLT). The mean-flow effects are usually described in terms of the divergence of the Eliassen–Palm flux (EPF) that drives a residual circulation [see Andrews et al. (1987)]. The underlying theoretical framework is known as the transformed Eulerian mean (TEM) equations. However, the TEM equations do not include the energy deposition that accompanies the EPF divergence.

The energy deposition is related to thermodynamic irreversibility. In the basic hydrodynamic equations of motion, this irreversibility is expressed in terms of mechanical and thermal dissipation due to molecular viscosity and heat conduction. In the troposphere, the dissipation takes place at the end of macroturbulent cascades of kinetic energy and available potential energy (e.g., Augier and Lindborg 2013; Brune and Becker 2013). These cascades are maintained by the generation of available potential energy due to differential heating (radiative, latent, and sensible heating). The mechanical dissipation (i.e., the frictional heating due to molecular viscosity) is of fundamental importance in the Lorenz energy cycle. As stated by Lorenz (1967), the frictional heating is the net diabatic heating of the atmosphere in climatological equilibrium.

In the middle atmosphere, turbulent energy cascades are induced by the breakdown of gravity waves (GWs) that propagate upward from the troposphere, leading to mechanical and thermal dissipation (e.g., Lindzen 1981; Hines 1997; Lübken 1997; Fritts and Alexander 2003). On the other hand, the convergence of the vertical potential energy flux of GWs gives rise to the energy deposition that...
contributes to the large-scale heat budget. The relation between the dissipation rates and the energy deposition is not straightforward. This holds not only for GWs but also for large-scale waves that may account for a significant vertical energy transport. The main purpose of the present paper is to derive an expression for the energy deposition of waves resolved in middle-atmosphere climate models and to estimate the importance of the energy deposition of thermal tides.

It is well known that the energy deposition of GWs yields a substantial contribution to the large-scale heat budget of the MLT (e.g., Fomichev et al. 2002). This contribution is said to be direct since it is not related to adiabatic heating or cooling by the residual circulation nor to the downward heat flux associated with GW damping. The correct formulation of the energy deposition was first mentioned by Hines and Reddy (1967). Estimates for direct GW heating were first given by Liu (2000). Further accounts of the direct heating of GWs can be found in several studies (e.g., Becker 2004; Akmaev 2007; Becker and McLandress 2009; Yigit et al. 2009; Shaw and Becker 2011). Using geometric height as the vertical coordinate, a main result is that the energy deposition is given by the convergence of the vertical pressure flux, which is positive in the case of wave breaking, plus an additional term given by the negative momentum flux times the wind shear. Hence, the energy deposition can be expressed in terms of wave fluxes and the mean flow. It is thus tempting to assume that the energy deposition is automatically simulated in a general circulation model (GCM) with regard to waves that are explicitly resolved. In the present study we show that this assumption is not valid and that the energy deposition of resolved waves is mainly due to the frictional heating associated with the momentum diffusion that damps the waves. It is therefore essential to include the complete frictional heating in a GCM in order to simulate the heat budget of the MLT correctly. This frictional heating can be represented in a GCM by means of the shear production associated with the employed macroturbulent diffusion scheme. Assuming a quasi-stationary turbulent kinetic energy equation, the shear production equals the molecular frictional heating minus the adiabatic conversion that takes place at the turbulent scale (e.g., van Mieghem 1973). In addition, molecular diffusion and ion drag become important for wave damping in the thermosphere (Hong and Lindzen 1976; Vadas and Fritts 2005; Vadas and Liu 2009), and corresponding frictional heating rates have to be considered as well.

Atmospheric waves contributing to the wave spectrum in the MLT other than nonorographic GWs and tides usually do not propagate over the entire altitude range. For example, quasi-stationary planetary Rossby waves and orographic GWs forced in the troposphere propagate only up to approximately the lower mesosphere in winter and drive the residual circulation in the stratosphere [see the review of Alexander et al. (2010)]. Traveling planetary waves in the MLT, on the other hand, are generated in situ by dynamical instabilities and generate an EPF divergence that opposes the GW drag (Plumb 1983; Norton and Thuburn 1999; McLandress et al. 2006; Pendlebury 2012). There are also stationary wave structures in the MLT, especially in the northern winter hemisphere. These are known to be caused at least partially by regional differences in refraction/filtering of GWs at lower altitudes and the resulting longitude-dependent GW drag in the MLT (Smith 2003; Lieberman et al. 2013). As will be discussed in section 2, quasigeostrophic waves are not expected to yield a significant energy deposition in the MLT.

The morphology of thermal tides has frequently been studied via observation and modeling. Reviews can be found in Forbes (2007) or Smith (2012), for example. The linear theory of tides (Chapman and Lindzen 1969) and linear numerical models that include specific damping mechanism for tides (e.g., Hong and Lindzen 1976) are usual tools to interpret tidal observations or the results from comprehensive GCMs. Linear numerical models of tides usually also include nontrivial mean flows (e.g., Grier et al. 2004).

Overall, the refraction, breakdown, and damping of GWs in the MLT are subject to a regular daily cycle that is induced by the tidal variations of the large-scale winds. Such GW effects then can give rise to a modulation of the tides. The role of GWs in modulating the phases and amplitudes of tides has been the subject of many model studies (e.g., Lu and Fritts 1993; Ortlund and Alexander 2006; Liu et al. 2008, 2014). It is typically found that the GW drag alters the phases and vertical wavelengths of the tides, whereas the GW-induced turbulent vertical diffusion gives rise to the damping of the tides. As pointed out by Ortlund and Alexander (2006), the simulated effects of GWs on the tides are sensitive to the details of how the GWs are represented. Therefore, a general consensus of how the morphology of tides is modulated by GWs has not yet been reached. In particular, model estimates usually involve GW parameterizations that are based on the conventional framework of quasi-stationary and single-column dynamics for the GWs. As noted by Senf and Achatz (2011), this approach may lead to significant deviations for the tidal amplitudes and phases when compared to the corresponding results from a three-dimensional and transient ray-tracing model. Further complications arise from wave–wave interactions among different tidal modes, as well as among tides and traveling planetary waves (e.g., Norton and Thuburn 1999; Moudden and Forbes 2014).

The question of how tides modulate the temporally averaged zonal-mean GW drag was investigated by Miyahara and Wu (1989). They employed an idealized
GCM where the zonal-mean flow was considered as the basic flow in the parameterization of GWs. They found that tides caused a westward zonal-wind anomaly in the mesopause region and in the lower thermosphere that was related to the westward EPF divergence of the tides. This wind anomaly then shifted the eastward GW drag in the summer hemisphere to higher altitudes, and the drag per unit mass became significantly stronger. Becker (2012) used a high-resolution mechanistic GCM with simplistic tidal forcing and resolved GWs. He found that, contrary to the result of Miyahara and Wu (1989), tides gave rise to a downward shift and weakening (per unit mass) of the GW drag in the summer MLT. Becker (2012) interpreted the downward shift as a result of the nonlinear dependence of GW instability on the tidal wind variations.

In this paper we use a mechanistic GCM to investigate the role of tides on the general circulation of the MLT. As usual, GWs are parameterized and evaluated at each model gridpoint and time step. We focus on the energy deposition of the tides and its importance for the zonal-mean sensible heat budget of the MLT, and we provide estimates of the generalized EPF divergence of the tides. In addition, we perform sensitivity experiments in order to estimate the zonal-mean model response to 1) the thermal excitation of tides and 2) to the energy deposition owing to parameterized GWs and resolved waves. Results are based on simulations using an updated version of the Kühnungen Mechanistic General Circulation Model (KMCM) as described in Becker et al. (2015). This model includes explicit computations of radiation and the tropospheric moisture cycle, as well as energetically consistent parameterizations of GWs, turbulent diffusion, and ion drag.

In section 2, we revisit the energetics of GWs using an anelastic version of the primitive equations, which are derived in the appendix, and we derive an expression for the energy deposition owing to resolved waves. A description of the KMCM and our simulations is given in section 3. In section 4, we present KMCM results regarding the contributions of the tides to the zonal-mean momentum and heat budgets of the MLT, and we discuss the model response in the aforementioned sensitivity experiments. Our conclusions are summarized in section 5.

2. Energy deposition: Parameterized gravity waves versus resolved waves

GCMs are usually based on the primitive equations with enthalpy (sensible heat) as the prognostic thermodynamic variable. This holds also for the KMCM. Therefore, in order to interpret model results with regard to the energy deposition, we must analyze the enthalpy budget. For this purpose we shall use a formalism that is analogous to the derivation of the energy deposition owing to GWs (e.g., Becker 2004). When using the primitive equations with pressure as vertical coordinate for this analysis, the subgrid-scale terms look cumbersome, and the notation cannot easily be reconciled with the notation commonly used in GW theory. Alternatively, when using $z$ as vertical coordinate, the averaging over scales is much more complicated; it requires density-weighted averages and an incompressibility approximation to exclude sound waves. Therefore, in order to keep the mathematical procedure as simple as possible, we utilize the primitive equations in the anelastic approximation.

The anelastic approximation invokes a hydrostatic reference state that depends on height $z$ only. Thermodynamic variables are then expanded as $X = X_0(z) + X_1(\lambda, \phi, z, t)$, where $X$ represents either density $\rho$, pressure $p$, temperature $T$, or potential temperature $\Theta$. Furthermore, $\lambda$ is longitude, $\phi$ is latitude, and $t$ is time. Since there are different kinds of anelastic and pseudo-incompressible equations in the literature (e.g., Durran 1989), and since these equations use the Exner pressure as prognostic thermodynamic variable, we derive in the appendix a version of the anelastic primitive equations that includes the enthalpy equation [see Eqs. (A19)–(A22)]. In this section, we extend these equations by external diabatic heating, as well as by subgrid-scale closures for ion drag and diffusion of momentum and sensible heat, and we average over the scale of GWs in the single-column approximation. This yields the following forced-dissipative anelastic primitive equations:

$$
\partial_t \mathbf{v} = \nabla \times (f + \xi)\mathbf{e}_z - \omega \partial_z \mathbf{v} - \frac{\nabla p}{\rho} + \frac{1}{\rho} \partial_z [\rho_r (K_z + v) \partial_z \mathbf{v} - \rho_r \mathbf{v} \cdot \mathbf{w}] + \nabla (K_h S_h) - \mathbf{D}_v,
$$

(1)

$$
0 = \nabla \cdot \mathbf{v} + \rho^{-1} \partial_t (\rho v), \quad \text{and}
$$

(3)

where $c_p$, $d$, $T$, $T_r$, $Q_{\text{rad}}$, $Q_{\text{lat}}$, $\partial$, $\mathbf{D}_v$, $S_h$, $K_h$, $K_z + v$, $\mathbf{e}_z$, and $\mathbf{w}$ are the specific heat at constant pressure, the coefficient of thermal compression, the temperature at the top of the model atmosphere, the radiation and latitudinal heating rates, the horizontal divergence, the horizontal wind, the horizontal eddy viscosity, and the pressure, respectively.
The nomenclature is defined in the appendix; v and \( \nabla \) denote the horizontal velocity field and nabla operator, \( \xi \) is the relative horizontal vorticity, \( f \) is the Coriolis parameter, and \( w \) and \( \mathbf{e}_z \) denote the vertical velocity and unit vector in the vertical direction. The GW perturbations are denoted by primes, and the average over the GW scale is indicated by an overline for the vertical fluxes of horizontal momentum and temperature in Eqs. (1) and (4), \( \overline{\mathbf{v} w} \) and \( \overline{T w} \), as well as for the frictional heating rates in the last row of Eq. (4). All other flow variables are assumed to be averaged over the GW scale.

The vertical momentum diffusion in Eq. (1) is standard and involves the turbulent and molecular diffusion coefficients \( K_z \) and \( \nu \). This term is written along with the momentum deposition due to GWs. The horizontal momentum diffusion in Eq. (1) is formulated with a stress tensor; that is, \( \mathbf{S}_h \) is the horizontal strain tensor and \( K_h \) the turbulent horizontal diffusion coefficient [see Becker and Burkhardt (2007), their Eqs. (7) and (8)]. The tensor \( \mathbf{D} \) describes the damping rate coefficients for ion drag [see Liu and Yeh (1969)]. The diabatic heating rates in Eq. (4) include radiative and latent heating, \( Q_{\text{rad}} \) and \( Q_{\text{lat}} \), respectively, as well as the vertical and horizontal diffusion of heat (last two terms in the first row), where the former is written along with the convergence of the vertical GW heat flux. The turbulent diffusion coefficients are scaled by a Prandtl number \( P_r \). The second row in Eq. (4) represents the frictional heating due to momentum diffusion and ion drag, as well as the negative buoyancy production of GW kinetic energy.

In accordance with the usual formulation of subgrid-scale closures for diffusion and GWs in GCMs, we have assumed that the diffusion coefficients are mean-flow quantities. Furthermore, we have utilized the incompressibility assumption—that is, \( \rho \rho_v = -\partial / \partial z = -T / T_r \) [see Eqs. (A5) and (A8)]—to rewrite the adiabatic heating term. Since the frictional heating is quadratic in the horizontal velocity field, it includes contributions from the resolved flow and from the GW perturbations. All frictional heating rates specified in Eq. (4) are positive definite, as is required by the second law of thermodynamics.\(^1\)

Note that ion drag does not represent an internal force like stress divergences due to pressure and momentum fluxes. Additionally, ion drag is linear in the flow velocity, and the associated rate of change of kinetic energy is always negative at each point in the flow. The same holds when the Lorentz deflection is included in the tensor \( \mathbf{D} \). Following Schmidt et al. (2006), we assume that the kinetic energy transferred to the plasma is thermalized locally so that the frictional heating due to ion drag is given by the negative rate of change of the kinetic energy.\(^2\)

As mentioned in section 1, the turbulent frictional heating mainly represents the molecular frictional heating in the real atmosphere, which takes place at the microscale and results from wave breaking and the so-induced forward kinetic energy cascades. In the thermosphere, the molecular kinematic viscosity becomes increasingly important with altitude and gives rise to damping of GWs (Vadas and Fritts 2005; Vadas and Liu 2009). The turbulent vertical diffusion coefficient is therefore completed by the kinematic molecular viscosity. The corresponding molecular heat conduction is included in terms of diffusion of temperature instead of potential temperature.\(^3\)

The system of Eqs. (1)–(4) conserves energy regarding the dynamics of the mean flow (see appendix). Energy conservation also holds regarding diffusion of the mean flow. Here, we have assumed a diagnostic turbulence model—that is, a quasi-stationary turbulent kinetic energy equation [see section 3 in Becker (2004)]. The system is, however, incomplete regarding the nonlinear GW terms. The reason is that the GW kinetic energy is another prognostic variable that must be evaluated. To formulate the GW kinetic energy equation, we add momentum diffusion and ion drag to the momentum equation [Eq. (A19)] and linearize about a mean flow. We then multiply by \( \mathbf{v} \) and average over the GW scale. Since the diffusion tendencies are assumed to be small, the work performed by friction on a control volume of the size of the GW scale is negligible, and we obtain the following approximations:

\[
\rho_v^{-1} \mathbf{v} \cdot \partial_z \left[ \rho_v (K_z + \nu) \partial_z \mathbf{v} \right] = -(K_z + \nu) \partial_z \mathbf{v}^2 \quad \text{and} \quad \overline{\mathbf{v} \cdot \nabla (K_h \mathbf{S}_h)} = -K_h \overline{|\mathbf{S}_h|^2}.
\]  

---

\(^1\) Once the (symmetric) stress tensor for momentum diffusion is given, the associated frictional heating (also known as the shear production) follows from the energy conservation law for arbitrary fluid volumes [see text of Lindzen (1990)]. Other forms of the frictional heating are sometimes used in climate models. But these are neither consistent with the energy conservation law nor with the second law [see discussions in Becker (2003) and Becker and Burkhardt (2007)].

\(^2\) Ion drag does not perform any net work on a control volume of the flow. This is different for internal forces, where the work performed is equal to the surface integral of the stress tensor times the velocity field.

\(^3\) Gassmann and Herzog (2015) recently discussed that the turbulent vertical diffusion of \( \Theta \) may lead to negative thermal dissipation. This issue is potentially important for the heat budget of the MLT. However, it is beyond the scope of the present study to further address this problem.
Using the single-column approximation and the hydrostatic and continuity equations for the GWs, the GW kinetic energy equation can then be written as

\[
\begin{align*}
\left( \partial_t + w \partial_z \right) \bar{\mathbf{v}}/2 &= - (\bar{\mathbf{v}} \cdot \partial_z \mathbf{v} - \rho_r^{-1} \partial_z \bar{\mathbf{p}} \cdot \mathbf{w} + g T_r^{-1} \mathbf{T} \cdot \mathbf{w}) \\
&\quad - (K_z + \nu) (\partial_z \mathbf{v})^2 - K_h |S_h|^2 + \mathbf{v} \cdot (D \mathbf{v}).
\end{align*}
\]

An assumption implicitly made in GW parameterizations is that the left-hand side (LHS) of Eq. (7) equals zero. In this quasi-stationary case, the frictional heating and buoyancy terms in Eq. (4) can be eliminated via

\[
E = - \rho_r^{-1} \partial_z \bar{\mathbf{p}} \cdot \mathbf{w} - (\bar{\mathbf{v}} \cdot \partial_z \mathbf{v}) - g T_r^{-1} \mathbf{T} \cdot \mathbf{w} + (K_z + \nu) (\partial_z \mathbf{v})^2 + K_h |S_h|^2 + \mathbf{v} \cdot (D \mathbf{v}).
\]

Here, \( E \) represents the well-known energy deposition of GWs (e.g., Hines 1997). Equation (8) shows that the energy deposition can be expressed in terms of the GW kinetic energy minus the buoyancy production. The buoyancy production in Eq. (7) (i.e., \( g T_r^{-1} \mathbf{T} \cdot \mathbf{w} = \theta_r \partial_z \mathbf{T} \cdot \mathbf{w} \)) is usually negative in regions of nonconservative wave propagation. As a result, each term on the RHS of Eq. (8) is positive. Note that we retain the horizontal diffusion terms for the GW dynamics. These terms are usually ignored in GW parameterizations but are important in GW-resolving simulations (Becker 2009). For the sake of completeness we also consider the effect of ion drag on the GWs. Inserting Eq. (8) into Eq. (4) leads to

\[
\begin{align*}
c_p \delta T &= - w g (1 + \bar{T}/T_r) + Q_{\text{rad}} + Q_{\text{lat}} \\
&\quad + \frac{c_p}{\rho_r} \left[ \rho_r \left( \frac{K_z + \nu}{P_r} \partial_z \mathbf{T} + \nu \partial_z T - \mathbf{T} \mathbf{w} \right) \right] \\
&\quad + c_p \nabla \cdot \left( \frac{K_h}{P_r} \nabla T \right) + E + (K_z + \nu) (\partial_z \mathbf{v})^2 \\
&\quad + K_h |S_h|^2 + \mathbf{v} \cdot (D \mathbf{v}),
\end{align*}
\]

which corresponds to the thermodynamic equation for a GCM with energetically consistent parameterizations of GWs, turbulent and molecular diffusion, and ion drag.

According to Eq. (8), the energy deposition primarily equals the dissipation of kinetic energy by momentum diffusion and ion drag. The remaining buoyancy flux term, on the other hand, can be related to the dissipation of GW potential energy. This is readily seen from the GW potential energy equation. To derive this equation, we linearize Eq. (A22) about a mean flow and add vertical and horizontal diffusion of heat (in terms of the temperature perturbation \( T \)). We then multiply by \( g^2 T/\left( N^2 T^2_r \right) \), where \( N \) is the buoyancy frequency of the reference state, and we average over the GW scale while noting that \( T_r \) and \( N \) vary much more slowly with height than \( \rho_r \) or \( \Theta_r \). The assumption that the diffusive GW heat fluxes are negligible across the GW scale leads to approximations that are analogous to Eqs. (5) and (6):
vertical momentum flux \(\text{(Miyahara and Wu 1989)}\), we anticipate that the convergence of the vertical pressure flux owing to tides is significant and will produce a significant energy deposition at high altitudes. On the other hand, waves that can approximately be described by quasi-geostrophic theory possess no significant vertical momentum flux and will therefore not produce a relevant energy deposition in the MLT.

From Eqs. (1)–(3) and (9) it is straightforward to write down the zonally averaged primitive equations in the TEM framework with the complete enthalpy equation. We use the notation of Lorenz (1967) and denote zonal means by brackets and deviations from the zonal mean by an asterisk. The residual circulation in the anelastic case is defined as

\[
v_{\text{res}} = [v] - \frac{1}{\rho_r} \partial_z \left( \rho_c \frac{[T^s u^s]}{g/c_p + \partial_z T} \right) \quad \text{and} \quad (13)
\]

\[
w_{\text{res}} = [w] + \frac{\partial \phi}{a \cos \phi} \left( \cos \phi \frac{[T^s u^s]}{g/c_p + \partial_z T} \right).
\]

The transformed zonal-mean zonal momentum equation and EPF components then are

\[
c_p \partial_z [T] = -c_p v_{\text{res}} \frac{\partial [T]}{a [T]} - c_p w_{\text{res}} \left( \frac{g}{a} + \partial_z [T] \right) + [Q_{\text{rad}}] + [Q_{\text{lat}}] - \frac{1}{\rho_r} \partial_z \left( \rho_c \frac{[T^s u^s]}{g/c_p + \partial_z T} \right) \frac{\partial \phi}{a [T]} + \rho_r [T^s w^s]
\]

\[+
\frac{1}{\rho_r} \partial_z \left( c_p \rho_r \left( \frac{T^o}{\Theta_r} \frac{K_z}{P_r} + \frac{[u^s T]}{P_r} - [T^s w^s] \right) \right) + \frac{c_p}{P_r} [\nabla \cdot (K_h \nabla T)]
\]

\[+
[E] - \frac{g}{T_r} [T^s w^s] + ((K_z + \nu)(\partial_z v)^2) + [K_h |S_h|^2] + [v \cdot (Dv)].
\]

Equations (13)–(17) are equivalent to the corresponding TEM equations given in [Andrews et al. (1987), their Eqs. (3.5.1), (3.5.3), (3.5.2a), and (3.5.2c)], except for the inclusion of GW effects, heat diffusion, and frictional heating. Applying the interpretation of the energetics of GWs to the TEM equations, we can conclude that the last row of Eq. (17) represents the complete zonal-mean energy deposition that is due to parameterized GWs and the resolved flow. While the buoyancy flux term [second term in the last row of Eq. (17)] is implicitly included in Eq. (3.5.2c) of Andrews et al. (1987), who used potential temperature instead of enthalpy as prognostic variable, the energy deposition was not considered.

In practice, we use pressure as vertical coordinate when computing the energy deposition due to the resolved flow. Assuming that resolved waves rather than the zonal-mean flow contribute to the frictional heating in the middle atmosphere, the energy deposition due to resolved waves can be computed using

\[
E_{\text{rw}} = \frac{R}{p} [T^s \omega^s] + [(K_z + \nu)(g p a^2 \omega)^2] + [K_h |S_h|^2] + [v \cdot (Dv)],
\]

where the subscript rw indicates resolved waves. The negative buoyancy production term has been reformulated according to Shaw and Becker (2011), where \(\omega\) represents the pressure velocity. Furthermore, the vertical wind shear has been rewritten according to the hydrostatic formula and the horizontal diffusion and dissipation is evaluated at surfaces of constant pressure [see Becker and Burkhardt (2007)]. The energy deposition given by Eq. (18) relates to

---

4 The quasi-geostrophic version was given in Becker (2012).

5 The zonally averaged hydrostatic, continuity, and gradient-wind equations are straightforward and not further specified here.

6 It is readily shown that \([T^s u^s]/(g/c_p + \partial_z T) = [\Theta u^s]/\partial_z [\Theta]\) holds in the anelastic approximation.
the tides if we insert just the tidal wind field and tidal temperature perturbation.

3. Model description

We use an updated version of KMCM as described in Becker et al. (2015, hereafter BKL15). In the following we describe the main features of this model.

The KMCM has a standard spectral dynamical core combined with a hybrid vertical coordinate (Simmons and Burridge 1981). We apply the same horizontal resolution as in BKL15—that is, triangular spectral truncation at wavenumber 32, but an extended model top, using 80 full model layers up to 8 $\times$ $10^{-5}$ hPa ($\sim$200 km). The model is equipped with an idealized radiation scheme, as well as with an explicit tropospheric moisture cycle, where the method of Schlutow et al. (2014) is used to compute the transport of water vapor. Land–sea contrasts are taken into account by including orography, land–sea masks for several surface parameters (albedo, relative humidity, heat capacity, and roughness length), and a simple slab ocean in order to close the surface energy budget and the radiation budget.

Parameterizations of GWs consist of the McFarlane scheme for orographic GWs (McFarlane 1987) and an extended Doppler-spread parameterization (DSP) for nonorographic GWs (Becker and McLandress 2009) that is based on the ideas of Hines (1997). The nonorographic GWs are launched isotropically in eight equidistant azimuthal directions at about 600 hPa. The following set of tunable parameters are used [for definition of symbols, see Becker and McLandress (2009)]: the minimum vertical wavenumber is $m_0 = 5 \times 10^{-4}$ m$^{-1}$; the maximum of the Desaubies spectrum at the launch level is set to a vertical wavenumber of $m^*_l = 4.2 \times 10^{-2}$ m$^{-1}$; the total squared horizontal wind variance at the launch level is $s_h = 0.87$ m$^2$ s$^{-2}$ between 35°S and 35°N, and $s_h$ decreases to 0.76 m$^2$ s$^{-2}$ toward the poles; the horizontal wavelength is 380 km; and the fudge factors are $\phi_1 = 1.5$, $\phi_2 = 0.3$, and $\phi_0 = 0.35$.

The KMCM employs a standard local vertical diffusion scheme (Holtslag and Boville 1993). Additional diffusion coefficients result from the GW schemes and from molecular diffusion in the thermosphere. To ensure a consistent scale interaction between parameterized GWs and the resolved flow, the complete vertical diffusion coefficient is used to damp parameterized GWs below their critical levels. In particular, the nonorographic GWs are strongly damped in the thermosphere by molecular diffusion; they have no notable effect on the circulation above about 10$^{-2}$ hPa ($\sim$110 km). This method is also applied in the Whole Atmosphere Community Climate Model (WACCM) (Garcia et al. 2007, their appendix A) in the context of a Lindzen-type parameterization. The solid gray line in Fig. 1a shows the prescribed asymptotic mixing length above 15 km for local vertical diffusion scheme. The solid black curve in the same panel gives the corresponding vertical diffusion coefficient.
coefficient in the climatological horizontal mean from the control simulation that is further described below. Figure 1b displays the corresponding vertical diffusion coefficients due to molecular viscosity (gray curve) and to turbulence generated by the breakdown of parameterized GWs (black curve). The vertical momentum diffusion of the resolved flow is completed by the associated frictional heating according to an energy preserving vertical discretization (Becker 2003; Boville and Bretherton 2003).

GCMs generally require some scale-selective damping of the horizontal wind in order to balance the horizontal energy and enstrophy cascades associated with the resolved flow. The KMCM employs a harmonic momentum diffusion for this purpose. The scheme is based on a symmetric stress tensor formulation and includes the frictional heating. The horizontal diffusion coefficient consists of a Smagorinsky-type (nonlinear) part and an additional linear part. Detailed information about the computation of the horizontal diffusion terms is given in Becker and Burkhardt [2007, their Eqs. (13), (15), (17)–(23), and (31)]. The free parameter of the nonlinear part [Eq. (13) in Becker and Burkhardt (2007)] is the horizontal mixing length as a function of height. The profile applied in the present model is given by the gray curve in Fig. 1c. The resulting Smagorinsky-type horizontal diffusion coefficient is displayed by the solid black curve (climatological horizontal-mean result from the control simulation). The additional prescribed horizontal diffusion coefficient (dashed curve in Fig. 1c) is significant above about 10^{-5} hPa and provides a sponge layer in the upper model domain.

The complete momentum diffusion applied in the KMCM preserves angular momentum and energy for arbitrary fluid volumes. In particular, the vertical energy flux and EPF by resolved waves are precisely deposited when these waves are damped by diffusion. This holds also for the sponge layer as formulated in the KMCM. In contrast, angular momentum and energy conservation are strongly violated when Rayleigh friction is used as a sponge layer.

To properly describe the damping of tides in the thermosphere, we also include ion drag and the associated frictional heating. Ion drag is particularly efficient in damping large-scale waves that have periods larger than about 1 h and propagate perpendicular to the direction of Earth’s magnetic field. Here we use the same formulation as in Fomichev et al. (2002; their section 3.3); that is, we consider only the damping of the horizontal wind and neglect the Lorentz deflection, we take the vertical profile for the damping coefficient from Hong and Lindzen (1976) for solar minimum conditions, and we use a simple formula for the dip angle. To account for the changes in the molecular composition above the mesopause, the current version of the KMCM applies alitudinal profiles for the heat capacity, gas constant, and molecular viscosity according to Vadas (2007).

Model results are computed from a 15-yr control simulation (after equilibration of climatology). Snapshots of the spectral state vector are archived every 90 min. All model quantities are reconstructed from these data. The tidal variations for a particular month are computed in the following way: All saved spectral state vectors of the 15 realizations of that month are rearranged such as to combine them into a single daily cycle; this model day then consists of 16 universal times (i.e., 1.30 a.m., 3.0 a.m., . . . , 24.0 p.m.), where each daytime contains 15 (number of years) times 30 (number of days for each month) snapshots. Taking the average over these 450 snapshots for each daytime gives a mean daily cycle representative for the chosen month. The mean daily cycle should include only the zonal-mean state, stationary waves, and tides. Because of the final number of snapshots that are averaged for each daytime, other wave components may also contribute to the diagnosed tides—for example, higher harmonics of the quasi-2-day wave. For the purpose of the present study, however, our simplified diagnostics of the total tidal wave field is considered to be sufficient.

Figure 2 illustrates the zonal-mean climatology for July and January of the control simulation in terms of the temperature along with the residual mass streamfunction and the ion drag (ID) and the zonal wind along with the sum of the resolved EPF divergence and the parameterized GW drag (GWD). The simulated wind and temperature fields are very reasonable, and the wave driving in the MLT compares well with other models (e.g., Richter et al. 2008). In particular, the model simulates a cold summer mesopause and a warm winter polar stratosopause as a consequence of the GW-driven residual circulation in the upper mesosphere. Figure 2 furthermore confirms that the hemispheric differences due to stronger quasi-stationary Rossby waves in the northern winter stratosphere are well captured. This becomes evident when comparing the polar night jets in Figs. 2b and 2d. As a result of this asymmetry, the winter westward GW drag in the MLT is clearly stronger in July than in January. As shown by BKL15 and Karlsson and Becker (2016), this leads to a colder summer polar mesopause during July than during January via the interhemispheric coupling mechanism. The zonal-wind structure in the MLT exhibits not only the wind reversals in the mesopause region but also additional reversals around 10^{-5} hPa (~150 km) in summer and 3 \times 10^{-5} hPa (~130 km) in winter, which compares well with results from comprehensive models (Fomichev et al. 2002; Schmidt et al. 2006).
It is important to validate the simulated tides. Figures 3a and 3b show the mean zonal-wind tides at 52°N for July and January after averaging the tides at the same local time over all longitudes. There is a predominant diurnal tide in the stratosphere and lower mesosphere, whereas a semidiurnal tide develops above about 0.03 hPa (70 km). This behavior is also reflected in Figs. 3c and 3d, which show the meridional-wind and temperature tides corresponding to Fig. 3a. The transition from a diurnal to a semidiurnal tide with altitude in the extratropics has been known also from other studies [see section 4.3 in Smith (2012)]. Figures 3a,c,d may be compared directly to ground-based observational results of Gerding et al. (2013, their Figs. 4 and 5) obtained at a single station in northern middle latitudes. Though the tidal amplitudes are somewhat smaller in the model, the phases fit reasonably well to the observational estimates. In particular, we expect that the simulated tidal variations of the zonal wind give rise to a modulation of the instantaneous GW drag. A more detailed analysis of the morphology of the tides simulated with the KMCM is beyond the scope of the present study.

To estimate the response of the MLT to tides, it is necessary to perform a sensitivity experiment where the excitation of tides is excluded. For this purpose, the radiative transfer calculations were performed as in the control run, but only the zonal means of the shortwave heating rates were fed into the thermodynamic equation of the model. In this way the excitation of thermal tides due to absorption of solar radiation was excluded while the zonal-mean radiative forcing of the model remained unchanged. The corresponding model experiment will be referred to as the no-tides simulation. A second sensitivity experiment compares the control simulation to a perturbation run where the complete energy deposition due to parameterized GWs and resolved waves was removed from the thermodynamic equation (no-energy-deposition run).

4. Results and discussion

In this section, we shall first discuss the mean-flow effects of tides as deduced from the control simulation.
This provides the basis for the interpretation of the model response to tides and energy deposition.

a. Mean-flow effects of thermal tides

Figures 4a and 4c show the EPF divergence of the resolved waves during July and January in the control simulation above 50 hPa (~20 km). The contribution from planetary waves (zonal wavenumbers 1–6 only) to the resolved EPF divergence is indicated by the black contours in Figs. 4b and 4d below 0.001 hPa (~90 km). While the upper troposphere (not shown) and the northern winter middle atmosphere exhibit the expected westward wave driving due to planetary Rossby waves, the summer mesopause region (i.e., around 0.003 hPa) confirms the well-known westward EPF divergence of in situ generated, westward traveling baroclinic waves (e.g., Norton and Thuburn 1996; Pendlebury 2012). Below that region, there is a significant positive EPF divergence that mainly reflects the resolved GW activity that is simulated even in a T32 model (McLandress et al. 2006). The model produces an eastward EPF divergence during July in the southern winter polar mesosphere. As shown by McLandress et al. (2006), this feature is due to eastward-traveling planetary waves that develop as a consequence of the reversed baroclinic instability in the southern winter lower mesosphere (strong temperature increase toward the pole), which is maintained by the strong westward GW drag aloft (cf. Fig. 2b).

The colors in Figs. 4b and 4d confirm former results of Miyahara et al. (1993) that the tides exhibit a negative EPF divergence in the lower thermosphere. In the present model simulation, the westward tidal wave drag maximizes at subtropical and middle latitudes around the lower edge of the sponge layer, reaching values of about −20 to −30 m s⁻¹ day⁻¹ during winter. There is an eastward tidal EPF divergence at polar latitudes in either hemisphere that indicates nonmigrating components with eastward phase propagation. The predominantly westward wave driving induces poleward circulation branches in either hemisphere above about 0.001 hPa (~90 km) as is visible in Figs. 2a and 2c. Figure 5 provides a more detailed picture of the residual circulation in the thermosphere, showing that the poleward circulation branches reach up to about 10⁻³ hPa (~150 km) in summer and up to the model top in winter. They are accompanied by upwelling in the mesopause region at low latitudes above 0.003 hPa (~85 km) and downwelling at high latitudes. The poleward circulation in the lower thermosphere in summer largely contributes to the rapid increase of the temperature with height above the mesopause. The poleward circulation in the winter lower thermosphere deviates from results obtained with the WACCM, which show a winter-to-summer pole residual circulation above the mesopause (Smith et al. 2011; Tan...
et al. 2012). A possible explanation for this difference is that the westward EPF divergence of the tides is stronger in the KMCM than in the WACCM. Also differences in the GW parameterization regarding GW components with high zonal phase speeds that can cause an eastward GW drag in the lower winter thermosphere may play a role.

In the summer hemisphere, the negative vertical wind shear that is induced by the tidal EPF divergence and by the radiatively determined state results into westward flow above about 10^{-5} hPa (see Figs. 2a and 2c). In winter, the westward tidal wave driving is not strong enough to prevent a positive vertical shear above about

---

**FIG. 4.** (a) EPF divergence due to resolved waves for July. (b) As in (a), but for the tides only. The black contours (drawn for ±2, 4, 8, 16, 32 m s^{-1} day^{-1}) display the EPF divergence due to planetary waves (zonal wavenumbers 1–6 only) up to 0.001 hPa. (b),(c) As in (a) and (b), respectively, but for January. The displayed altitude regime extends from about 20 to 200 km.

**FIG. 5.** Temperature (colors), residual mass streamfunction (back contours for ±0.001, 0.01, 0.1, 10 × 10^6 kg s^{-1}), and the sum of resolved EPF divergence, parameterized GWs, and ion drag (white contours for ±5, 10, 20, 40, 80, 120 m s^{-1} day^{-1}) for (a) July and (b) January. The displayed altitude regime extends from about 70 to 200 km.
$3 \times 10^{-4}$ hPa that is dominated by the radiatively determined state and leads to eastward flow above about $3 \times 10^{-5}$ hPa ($\sim 130$ km). The resulting zonal-wind structure in the upper model domain induces an ion drag that drives a summer-to-winter pole circulation (Fig. 5).

Figures 6a and 6c show the complete energy deposition in the MLT due to parameterized GWs and resolved waves for July and January. At middle latitudes and up to about 0.001 hPa, the model result confirms the well-known behavior for GWs: there is a sudden onset of the energy deposition around 0.01 hPa ($\sim 80$ km) in summer, reaching about 7 K day$^{-1}$ around 0.005 hPa for the present model tuning; on the other hand, the energy deposition in winter is significant throughout the mesosphere but generally smaller than during summer. Other examples for this summer–winter asymmetry can be found in, for example, Lübken (1997, his Figs. 7a and 8), Fomichev et al. (2002, their Fig. 8a), and Becker (2004, his Fig. 5). The energy deposition in Figs. 6a and 6c exhibits a significant heating in the thermosphere of several 10 K day$^{-1}$. Figures 6b and 6d reveal that this energy deposition is mainly due to the tides. Parameterized GWs do not notably contribute in the thermosphere for the present model; those GWs that are not yet obliterated around the mesopause are efficiently damped by molecular diffusion further above.

Figure 7a shows the horizontally and annually averaged energy deposition due to all waves and due to only the tides (black and gray solid curves), confirming that the latter give the main contribution above about 0.0003 hPa ($\sim 100$ km). The contributions from the linear horizontal momentum diffusion (dashed lines in Fig. 7a) reveal that the sponge layer becomes significant only above about $3 \times 10^{-6}$ hPa ($\sim 170$ km). Even within the sponge layer, the major contributions to the energy deposition are due to other processes. Figure 7b shows the individual contributions to the energy deposition of the resolved waves that result from ion drag, nonlinear horizontal diffusion (i.e., without the effect of the sponge layer), vertical diffusion, and the negative buoyancy production. Above about $3 \times 10^{-5}$ hPa ($\sim 130$ km), the frictional heating due to ion drag is the dominant contribution. The frictional heating due to nonlinear horizontal momentum diffusion dominates below that level. The frictional heating owing to vertical momentum diffusion (mainly due to molecular viscosity) and the negative buoyancy production (mainly induced by molecular heat diffusion) increase with altitude up to the onset of the sponge layer. These terms
are always less important than the contribution from either nonlinear horizontal diffusion or ion drag.

According to these model results, the energy deposition of the tides contributes significantly to the heat budget of the lower thermosphere. The estimate of a several $10 \text{ K day}^{-1}$ heating rate is quite significant, especially when compared to the cooling by molecular heat conduction which reaches about $-130 \text{ K day}^{-1}$ in the upper model domain (not shown).

b. Model response to the excitation of thermal tides

We now consider the first sensitivity experiment described in section 3—that is, the differences in the control run from the no-tides run where all shortwave radiative heating rates are substituted by their zonal means. Corresponding differences will henceforth be denoted as model response to tides. Figure 8 shows the response in the MLT for the zonal-mean temperature and zonal wind during July and in the annual mean. The climatology of the control run is shown as black contours for reference. The model response to tides is characterized by a substantial warming and westward acceleration in the lower thermosphere. While the overall warming in the altitude regime above about $10^{-3} \text{ hPa}$ ($\sim 90 \text{ km}$) is related to the energy deposition of the tides, the westward acceleration results from the predominantly westward tidal EPF divergence.

Figure 9 shows more details of the zonal-mean zonal forces and the energy deposition. Figure 9a depicts the response of the parameterized GW drag during July (with results from the control run being included by black contours). From about 0.01 to 0.0003 hPa (80 to 100 km) we can infer a reduction of the maximum GW drag in each hemisphere, which is accompanied by a slide amplification of the GW drag around 0.03 hPa (70 km). The details of this response may vary among models, depending on the morphology of the tides and the representation of GWs. Nevertheless, a qualitative explanation for a reduction and downward shift of the GW drag by the tides is straightforward (Becker 2012).

Let us consider the dispersion relation for midfrequency GWs in combination with the fact that mainly GWs with phase speeds of more than $10 \text{ m s}^{-1}$ contribute to the GW drag in the extratropical mesosphere (Lindzen 1981). These waves propagate nearly conservatively up to their breaking levels. When they are refracted to higher intrinsic frequencies during the westward (summer) or eastward (winter) phase of the zonal-wind tide, they still propagate nearly conservatively. The reversed phase of the tide, however, gives rise to their refraction to lower intrinsic frequencies and shorter vertical wavelengths such that the GWs encounter dynamic instability at somewhat lower altitudes. The net effect is a downward shift of the breaking levels and a reduction of the maximum zonal-mean GW drag per unit mass.

The temperature and wind responses shown in Figs. 8a and 8b are consistent with this notion. They are, however, superposed with the direct effects of the tides in the lower
thermosphere. Figure 9b shows the annual-mean response of the EPF divergence owing to resolved waves. Above the mesopause the pattern largely reflects the EPF divergence in the control run, which in turn is mainly due to the tides (see Fig. 4). In the annual mean, the thermospheric EPF divergence is somewhat stronger than during the solstices since the tidal amplitudes maximize around equinox. The poleward circulation branches in the lower thermosphere that are driven by the tidal EPF divergence lead to the structure of the temperature response that is visible in Figs. 8a and 8c between about 10^{-3} and 10^{-4} hPa (~90 and 110 km); that is, the strongest warming is found at polar latitudes. In turn, this temperature response is linked to the predominantly westward wind response visible in Figs. 8b and 8d, since the gradient–wind balance applies to the zonal-mean flow at least qualitatively in the MLT (McLandress et al. 2006). The westward zonal-wind response extends into the upper model domain and, hence, induces a substantial eastward response of the zonal force due to ion drag (Fig. 9c). In turn, the ion drag response induces equatorward circulation branches above about 10^{-5} hPa (~150 km) in the control run relative to the no-tides run (not shown). This explains why the meridional structure of the temperature response in Figs. 8a and 8c reverses above about 5 \times 10^{-5} hPa (~120 km) and is accompanied by a positive vertical shear in the zonal-wind response (Figs. 8b and 8d).

The response of the energy deposition in the lower thermosphere strongly resembles the energy deposition in the control run (Fig. 9d)—that is, a substantial contribution to the large-scale enthalpy budget of the thermosphere is missing in the no-tides run. The meridional distribution of the annual-mean energy deposition response shows maximum heating rates at middle and high latitudes in the thermosphere. This indicates that those tidal components that propagate to higher latitudes also propagate to higher altitudes as compared to other tidal components that are more confined to low latitudes. This is consistent with the notion that the diurnal tide, which dominates in the MLT at low latitudes, has shorter vertical wavelength and is thus more susceptible to damping by vertical diffusion than the semidiurnal tide, which dominates at higher latitudes [Smith (2012), her section 4.3.2; see also references therein].
As discussed in Zhu et al. (2010), Shaw and Becker (2011), and Zhu et al. (2011), there is no general consensus yet in the modeling community about the energy deposition of waves. Particularly the theory of the energetics of GWs as outlined in section 2 was questioned by Zhu et al. (2010, 2011). The importance of the energy deposition in the MLT, however, can be demonstrated by comparing the control run against the no-energy-deposition run in which the sum of energy deposition rates due to parameterized GWs and resolved waves,

$$E_{\text{rw}}$$ [see Eqs. (8) and (18)], was subtracted from the rhs of the thermodynamic equation of the model. Results for the temperature response during July and in the annual mean are shown in Fig. 10. The heating effect of the energy deposition is substantial in the MLT. The structure of the temperature response is similar to the response to tides (Fig. 8). A closer inspection of the model data showed that the energy deposition leads to lower static stability in the control run from the upper mesosphere on. As a result, the parameterized GW drag in the control run is strongly reduced and shifted to lower altitudes relative to the no-energy-deposition run. Stated otherwise, the GW drag in the no-energy-deposition run is very unrealistic; it is much too strong (more than 450 m s\(^{-1}\) day\(^{-1}\) in northern summer; not shown) and also too high in altitude. This explains part of the very strong and positive temperature response in the polar summer MLT around 3 × 10\(^{-5}\) hPa (∼100 km) in Fig. 10a. Another consequence of the energy deposition is a weaker GW-induced vertical diffusion coefficient. This gives rise to stronger tides and stronger poleward circulation branches in the lower thermosphere. The resulting westward wind response (not shown) induces an eastward response of the ion drag like in the previous sensitivity experiment. Hence, also the structure of the temperature response above about 3 × 10\(^{-5}\) hPa (∼130 km) is qualitatively similar in both cases.

5. Conclusions

We employed a mechanistic middle-atmosphere GCM with a model top around 8 × 10\(^{-6}\) hPa (∼200 km) in order to estimate the effects of thermal tides on the general circulation of the mesosphere–lower thermosphere (MLT).
The explicitly simulated dynamics (including tides and planetary waves) is comparable to the dynamics simulated with comprehensive middle-atmosphere GCMs. A specific property of the present model is that subgrid-scale parameterizations for gravity waves (GWs), momentum diffusion, and ion drag are formulated in an energy conserving way. In particular, the energy deposition related to the parameterized GW drag and the frictional heating (mechanical dissipation) that arises from momentum diffusion and ion drag are included. The sponge layer of the model is due to a linear harmonic horizontal diffusion that preserves angular momentum and energy. This sponge layer is only of secondary importance since the ion drag is very efficient to dissipate most of the wave energy in the upper model domain.

To derive an expression for the energy deposition of resolved waves, we first recapitulated the energetics of GWs in the single-column approximation for an anelastic version of the primitive equations. We emphasized that the usual formulation of the energy deposition in GW parameterizations is based on the assumption of a quasi-stationary energy equation for the GWs. When this assumption is not made, the energy deposition is given by the frictional heating minus the buoyancy production of GW kinetic energy. Following this insight, which was also mentioned in previous studies, zonal averaging of the thermodynamic equation gives rise to a corresponding expression that can be used to diagnose the energy deposition of waves resolved in a GCM [see Eq. (18)].

The energy deposition in the MLT is the irreversible heating that occurs when vertically propagating waves dissipate either as a result of wave breaking and turbulence or directly as a result of molecular viscosity and ion drag. The energy available for this process is generated in the region of wave excitation at lower altitudes—for example, typically in the troposphere in the case of GWs. Though the energy deposition mainly corresponds to the frictional heating, both terms are not equivalent. For example, the frictional heating in the troposphere is an essential part of the Lorenz energy cycle; it is maintained by buoyancy production of wave kinetic energy through baroclinic instability and differential heating. In contrast, the buoyancy production of wave kinetic energy is negative in the case of energy deposition, and wave kinetic energy is generated by the convergence of the vertical pressure flux (or potential energy flux). Hence, the energy deposition becomes relevant when vertical wave propagation is accompanied by a strong vertical pressure flux. Since this flux is proportional to the absolute vertical flux of horizontal momentum, quasigeostrophic waves hardly contribute to the energy transfer from low to high altitudes. This is, however, different for GWs and tides.

The climatology of the complete energy deposition as simulated with the present GCM confirms the well-known behavior for GW breakdown in the MLT, including the summer–winter asymmetry. In addition, there is substantial energy deposition in the lower thermosphere, where resolved waves are strongly damped by momentum diffusion and ion drag. This heating rate amounts to several tens of kelvins per day and is mainly caused by the tides.

We performed a sensitivity experiment with all radiative heating rates substituted by their zonal means such as to switch off the excitation of thermal tides. The difference between the control simulation and the model run without tides reveals the tidal effects on the general circulation. The difference in the lower thermosphere is characterized by a warming that is very pronounced in the polar regions. On the other hand, the energy deposition of the tides is significant at all latitudes and has weak maxima at middle and high latitudes. The redistribution of heat in the lower thermosphere is predominantly due to the westward Eliassen–Palm flux (EPF) divergence of the tides, which drives poleward circulation branches in both hemispheres of the lower thermosphere, with upwelling in the tropics and downwelling at polar latitudes.
At higher altitudes (above ~10^{-5} hPa corresponding to ~150 km), the westward zonal-wind response induces an eastward response of the ion drag. As a result, the latitudinal structure of the temperature response reverses at these heights, with minimum heating or even cooling at polar latitudes.

The model response to the excitation of tides also reveals a significant indirect effect on the general circulation in the upper mesosphere in terms of a downward shift of the breaking levels of parameterized GWs. A plausible interpretation is that the GWs become more dynamically unstable when the horizontal wind tide is in the direction of GW phase propagation, whereas the reduced instability in the opposite phase is less important. Indeed, the simulated zonal-wind tides in the summer MLT having amplitudes of ±10 m s^{-1} are supposed to induce such a modulation of the mean GW drag. This interpretation suggests that events of strong activity of westward-traveling waves in the summer mesosphere can also cause a reduction and downward shift of the GW drag.

The energy deposition of resolved waves is incomplete in GCMs when frictional heating is ignored or flawed. Though the contribution associated with molecular vertical momentum diffusion and ion drag is often included in GCMs extending into the thermosphere, it is the scale-selective damping of momentum by nonlinear horizontal diffusion that dissipates most of the energy of resolved waves in the MLT up to about 3 × 10^{-5} hPa (~130 km) in the present model. Conventional measures to parameterize this damping such as hyperdiffusion, numerical filtering, or Rayleigh friction are not consistent with angular momentum, energy conservation, and the second law of thermodynamics; they can hardly be used to describe the energy deposition of resolved waves. At altitudes above ~130 km, however, the ion drag dissipates most of the energy of resolved waves for the present model setup.

We demonstrated the importance of the energy deposition by parameterized and resolved waves by performing a second sensitivity experiment in which the energy deposition was neglected. We found that the energy deposition in the mesopause region causes the GWs to break at lower altitudes. As a result of the damping effect that GWs have on the tides in the present model, the tides are significantly stronger from the upper mesosphere on when the energy deposition is included. Therefore, the model response to the energy deposition is comparable to the model response to tides, though the maximum temperature response of more than 100 K above the summer mesopause is much stronger.

Summarizing, thermal tides have significant direct and indirect effects on the general circulation of the MLT that are not negligible. Particularly the EPF and energy deposited in the lower thermosphere are pivotal for the momentum and energy budgets in that region. Whereas the EPF divergence of the tides and the tidal modulation of the zonal-mean GW-drag have been analyzed in previous studies, the energy deposition has not been investigated until this present work. The mean-flow effects caused by thermal tides require further investigation based on comprehensive GCMs that 1) extend into the thermosphere, 2) employ consistent subgrid-scale diffusion schemes, and 3) include an improved representation of GW–tidal interactions. GW-resolving GCMs are expected to lead to promising new results in the future, since they account for GW intermittency and the generation of secondary GWs. Such GW effects were not accounted for in the present study but are expected to be quite important for the dynamics and variability in the thermosphere.

Acknowledgments. The author is grateful to S. L. Vadas and D. C. Fritts for inspiring discussions and comments on the manuscript. The comments by Jia Yue, two anonymous reviewers, and the editor Anne K. Smith are gratefully acknowledged. This study was partly funded by the Leibniz Society via the SAW project WaTiLa.

APPENDIX

An Anelastic Version of the Primitive Equations with Enthalpy as the Prognostic Variable

An anelastic version of the fluid dynamical equations of motion is advantageous in atmospheric applications when it is necessary to exclude sound waves. For example, the linear solutions for gravity waves are more easily obtained when using the anelastic equations. Different versions of anelastic or pseudo-incompressible approximations have been proposed in the literature. For an overview, the reader is referred to Durran (1989) and Achatz et al. (2010). Here, we follow the approach used by Vallice (2006, his chapter 2.5). We first consider the full Eulerian equations in Cartesian coordinates, from which we will then obtain the corresponding primitive equations.

As usual, we decompose density, pressure, temperature, and potential temperature as

\[
\rho = \rho_r(z) + \rho(x, y, z, t) \\
\rho = \rho_r(z) + \rho(x, y, z, t) \\
T = T_r(z) + T(x, y, z, t) \\
\Theta = \Theta_r(z) + \Theta(x, y, z, t),
\]

where the subscript \( r \) denotes a hydrostatic reference state that depends only on \( z \) and where the deviations are denoted by \( \rho, \rho, T, \) and \( \Theta, \) which depend on the spatial
coordinates \((x, y, z)\) and time \((t)\). The reference state fulfills the equation of state for an ideal gas with regard to the temperature and potential temperature; that is, \(RT_r = p_r/\rho_r\) and \(\Theta_r = T_r(p_{00}/p_r)^{R/c}\) with \(p_{00} = 1013\ \text{hPa}\). Here, \(R\) is the gas constant and \(c_p\) the heat capacity per unit mass at constant pressure.

Accordingly, the density velocity that performs work

\[
0 = \nabla \cdot (\rho_r \mathbf{v}) + \partial_z (\rho_r w),
\]

(A2)

where \(\nabla\) and \(\mathbf{v}\) denote the horizontal nabla operator and velocity field, respectively, and \(w\) is the vertical velocity. Accordingly, the density velocity that performs work becomes \(d\rho_r = wd\rho_r\), and an internal energy equation consistent with Eq. (A2) must be written as

\[
d_e = c_v dT = \frac{p_r + \hat{\rho}}{(p_r + \hat{\rho})} w d\rho_r,
\]

(A3)

where \(c_v\) denotes the heat capacity per unity mass at constant density. The second part of the anelastic approximation states that the perturbation quantities in Eq. (A1) are small as compared to the reference values and that the density perturbation only depends on the potential temperature perturbation. Hence the exact thermodynamic relation,

\[
\frac{d\Theta}{\Theta} = -\frac{d\rho}{\rho} + \frac{c_v}{c_p} \frac{dp}{\rho},
\]

(A4)

leads to the anelastic version,

\[
\frac{\hat{\Theta}}{\Theta_r} = -\frac{\hat{\rho}}{\rho_r},
\]

(A5)

since

\[
\frac{|\hat{\rho}|}{\rho_r} \gg \frac{|\hat{\rho}|}{\rho_r}.
\]

(A6)

These approximations allow the expansion of Eq. (A3) as

\[
d_e = \frac{p_r}{\rho_r} \left( 1 - 2 \frac{\hat{\rho}}{\rho_r} \right) w d\rho_r,
\]

(A7)

and they imply that

\[
\frac{\hat{\Theta}}{\Theta_r} = \frac{\hat{T}}{T_r}.
\]

(A8)

The anelastic version of the enthalpy equation is obtained by noting that

\[
h = e + \frac{p_r}{\rho_r} \left( 1 - \frac{\hat{\rho}}{\rho_r} \right),
\]

(A9)

where again the approximation Eq. (A6) has been used. When computing the total time derivative of Eq. (A9), we insert Eq. (A7) and differentiate the second term on the rhs of Eq. (A9) only with regard to the reference state. Furthermore, we make use of the hydrostatic relation for the reference state, \(d\rho_r = -g\rho_r\), where \(g\) is the acceleration due to gravity. The final result is

\[
d_t h = c_p d_t T = -wg \left( 1 - \frac{\hat{\rho}}{\rho_r} \right),
\]

(A10)

To derive the anelastic version of the horizontal Eulerian momentum equation,

\[
d_t \mathbf{v} = - (\rho_r + \hat{\rho})^{-1} \nabla (p_r + \hat{\rho}),
\]

(A11)

we simply neglect \(\hat{\rho}\), yielding

\[
d_t \mathbf{v} = -\rho_r^{-1} \nabla \hat{\rho}.
\]

(A12)

The anelastic approximation is less straightforward for the vertical momentum equation,

\[
d_t w = - (\rho_r + \hat{\rho})^{-1} \partial_z (p_r + \hat{\rho}) - g.
\]

(A13)

To preserve the buoyancy force, we first multiply Eq. (A13) by the density, \(\rho_r + \hat{\rho}\), and then neglect \(\hat{\rho}\) on the lhs of Eq. (A13). After dividing by \(\rho_r\), we obtain

\[
d_t w = -\rho_r^{-1} \partial_z \hat{\rho} - \rho_r^{-1} \hat{\rho} g
\]

(A14)

as a preliminary result. Combining Eqs. (A12) and (A14) with Eq. (A2) allows to derive the kinetic energy equation in flux form:

\[
d_t (\rho_r k) + \nabla \cdot (\mathbf{v} \rho_r k) + \partial_z (w \rho_r k) = -\mathbf{v} \cdot \nabla \hat{\rho} - w \partial_z \hat{\rho} - w g \rho_r, \]

(A15)

where \(k = (\mathbf{v}^2 + w^2)/2\). The flux form of the enthalpy equation, Eq. (A10), is

\[
\partial_t (\rho_r h) + \nabla \cdot (\mathbf{v} \rho_r h) + \partial_z (w \rho_r h) = -wg \rho_r \left( 1 - \frac{\hat{\rho}}{\rho_r} \right).
\]

(A16)

Adding Eqs. (A15) and (A16), the buoyancy terms proportional to \(\hat{\rho}\) cancel. Furthermore, the first term on the right-hand side (rhs) of Eq. (A16) can be written in a flux form using \(d_t \Theta = wg\). However, the pressure work on the rhs of Eq. (A15) cannot be written as a flux...
divergence. This becomes only possible if we specify the vertical momentum equation as

\[ \frac{d}{dt} \omega = -\frac{\partial}{\partial \rho_r} \left( \frac{\tilde{\rho}}{\rho_r} \right) - \frac{\tilde{\rho} g}{\rho_r}. \]  

(A17)

Hence, \( \tilde{\rho} \rho_r^{-2} \frac{d}{dt} \rho_r \approx -\rho_r^{-1} \tilde{\rho} g \) must be added to the rhs of Eq. (A14). This manipulation is consistent with the anelastic approximation according to Eq. (A6). It is required not only for energy conservation but also to recover the correct dispersion relation for GWs with long vertical wavelengths.

Summarizing, an anelastic version of the Eulerian fluid dynamical equations with enthalpy as prognostic thermodynamic variable consists of Eqs. (A12), (A17), (A2), and (A10). This set fulfills the following energy conservation law:

\[ \delta_t [\rho_r (k + h + \Phi)] + \nabla \cdot [\rho_r (k + h + \Phi)] + \delta_z [\rho_r (k + h + \Phi)] = \nabla \cdot (\tilde{\rho} \omega) - \partial_z (\tilde{\rho} \omega). \]  

(A18)

Furthermore, owing to the equations of state for the reference state, and owing to Eqs. (A5) and (A8), the enthalpy Eq. (A10) is equivalent to \( \delta_t \Theta = 0 \).

Deriving the primitive equations from the anelastic Eulerian equations is straightforward. We define the horizontal velocity field and nabla operator in spherical coordinates, invoke the shallow-atmosphere approximation, add the Coriolis force, and substitute the vertical momentum equation by the hydrostatic approximation. This leads to\(^1\)

\[ \delta_t v = \nabla \times \left( f + \xi \right) e_z - \omega \partial_z v - \nabla^2 \frac{\tilde{\rho} v}{2} - \tilde{\rho} \frac{\partial \omega}{\partial \rho_r}, \]  

(A19)

\[ \delta_t \frac{\tilde{\rho}}{\rho_r} = -g \frac{\tilde{\rho}}{\rho_r} \]  

(A20)

\[ 0 = \nabla \cdot v + \rho_r^{-1} \delta_z (\rho_r \omega), \]  

(A21)

\[ c_p \frac{d}{dt} T = -w g (1 - \tilde{\rho}/\rho_r), \]  

(A22)

where \( \xi \) is the horizontal vorticity and \( e_z \) is the unit vector in the vertical direction. Defining \( h_{\text{tot}} = c_p T + v^2/2 + \Phi \), the energy conservation law related to Eqs. (A19)–(A22) yields

\[ \delta_t (\rho_r h_{\text{tot}}) + \nabla \cdot (\omega_r h_{\text{tot}}) + \delta_z (\rho_r w h_{\text{tot}}) \]  

\[ = -\nabla \cdot (\tilde{\rho} \omega) - \delta_z (\tilde{\rho} \omega). \]  

(A23)

\(^1\)The anelastic primitive equations in Becker (2012) are not consistent with the current approach.

REFERENCES


