

Explicit Global Simulation of Gravity Waves in the Thermosphere

 Erich Becker^{1,2}  and Sharon L. Vadas³ 
¹Leibniz Institute of Atmospheric Physics, Kühlungsborn, Germany, ²Now at Northwest Research Associates, Boulder, CO, USA, ³Northwest Research Associates, Boulder, CO, USA

Key Points:

- GWs in winter thermosphere are mainly tertiary GWs generated in lower thermosphere
- GWs in winter thermosphere have large vertical wavelengths and propagate against background tidal winds
- Medium-small-scale GWs create hot spots in winter thermosphere like orographic hot spots in stratosphere

Correspondence to:

 E. Becker,
 erich.becker@nwra.com

Citation:

 Becker, E., & Vadas, S. L. (2020). Explicit global simulation of gravity waves in the thermosphere. *Journal of Geophysical Research: Space Physics*, 125, e2020JA028034. <https://doi.org/10.1029/2020JA028034>

Received 21 MAR 2020

Accepted 6 AUG 2020

Accepted article online 24 AUG 2020

Abstract We present a new version of the high-resolution Kühlungsborn Mechanistic general Circulation Model (KMCM) extended to $z \sim 450$ km. This model is called HIAMCM (HI Altitude Mechanistic general Circulation Model) and explicitly simulates gravity waves (GWs) down to horizontal wavelengths of $\lambda_h \sim 165$ km. We find predominant tertiary GWs in the winter thermosphere at middle/high latitudes. These GWs typically have horizontal wavelengths $\lambda_h \sim 300$ – $1,100$ km, ground-based periods ~ 25 – 90 min, and intrinsic horizontal phase speeds $c_{th} \sim 250$ – 350 m s⁻¹. Above $z \sim 200$ km, the predominant GW horizontal propagation directions are roughly against the background winds from the diurnal tide; the GWs propagate mainly poleward at midnight, eastward at 6 local time (LT), equatorward at noon, and westward at 18 LT. Wintertime GWs at $z \sim 300$ km having 165 km $\leq \lambda_h \leq 330$ km create a large hot spot over the Southern Andes/Antarctic Peninsula that agrees well with quiet time satellite measurements. Due to cancellation effects, the time-averaged zonal mean Eliassen-Palm flux divergence from the resolved GWs in the thermosphere is negligible compared to that of the tides and compared to the zonal component of the time-averaged zonal mean ion drag. We also find that the thermospheric GWs dissipate mainly from macroturbulent diffusion and, above $z \sim 200$ km, from molecular diffusion, whereas the tides dissipate mainly from ion drag. The averaged dissipative heating in the thermosphere due to tides is much stronger than that due to GWs.

1. Introduction

The global circulation in the mesosphere is mainly driven by the wave mean flow interaction due to internal gravity waves (GWs) (e.g., Smith, 2012). Further contributions result from in situ generated planetary waves (McLandress et al., 2006). The circulation in the lower thermosphere is strongly driven by thermal tides and by ion drag (Becker, 2017). It is commonly believed that the GWs relevant for the circulation and variability of the stratosphere and mesosphere are of tropospheric origin. That is, these are *primary* GWs generated by flow over orography, moist convection, and “spontaneous emission” from tropospheric jets and fronts (see reviews of Fritts & Alexander, 2003; Plougonven & Zhang, 2014). On the other hand, GWs can be generated in the stratosphere and lower mesosphere by the nonlinear dynamics of the polar vortex (e.g., Sato & Yoshiki, 2008; Sato et al., 2012; Shibuya et al., 2017; Triplett et al., 2017). Furthermore, secondary GW generation due to the intermittent body forces resulting from the dissipation of primary GWs is a predominant mechanism in the southern winter hemisphere over orography (Becker & Vadas, 2018; Vadas & Becker, 2018; Vadas et al., 2018). We refer to this latter process as the *body-force mechanism* (Fritts et al., 2006; Vadas et al., 2003). Many recent observation-based studies of GWs in the mesosphere and lower thermosphere (MLT) lend support to the applicability of the body-force mechanism (e.g., Bossert et al., 2017; Chen et al., 2013; de Wit et al., 2017; Huang et al., 2017; Vadas et al., 2018; Vargas et al., 2016; Zhao et al., 2017). In particular, the observation of persistent large-amplitude inertia GWs having large periods and large inferred spatial scales in the southern winter mesopause region over McMurdo Station in the Antarctic can be explained via this mechanism (Becker & Vadas, 2018; Chen & Chu, 2017; Chen et al., 2016; Vadas & Becker, 2018; Vadas et al., 2018). Furthermore, the eastward Eliassen-Palm flux (EPF) divergence of secondary GWs generated by the body-force mechanism reduces the downwelling over the winter pole around the mesopause or even gives rise to upwelling, thus contributing to cold temperatures and eastward winds in the winter polar mesopause region (Becker & Vadas, 2018).

When secondary GWs break and deposit their momentum (and energy), tertiary GWs are generated by the body-force mechanism. It is very likely that both secondary and tertiary GWs are significant in the

thermosphere (Vadas & Crowley, 2010, 2017; Vadas & Liu, 2009, 2013; Vadas & Nicolls, 2009; Vadas et al., 2014, 2019). Indeed, the characteristics of GWs in the thermosphere that induce traveling ionospheric disturbances (TIDs) (e.g., Hocke & Schlegel, 1996; Nicolls et al., 2014) often cannot be reconciled with these GWs having their sources below the mesopause (e.g., Vadas & Crowley, 2010). On the other hand, the body-force mechanism can explain such GW activity in the thermosphere (Vadas & Becker, 2019; Vadas & Liu, 2013; Vadas et al., 2019). It is only recently that our community is beginning to understand the importance of this multistep vertical coupling mechanism by which primary GW generation in the troposphere leads to dynamical control of the thermosphere/ionosphere via secondary and tertiary or higher-order GWs. For the rest of this paper, we refer to tertiary and higher-order GWs as tertiary GWs.

Recent satellite observations lend evidence to the importance of this mechanism. For example, according to the analysis of Trinh et al. (2018), there is a major traveling atmospheric disturbance (TAD) hot spot at $z \sim 250\text{--}450$ km over the Southern Andes/Antarctic Peninsula region during June to August. Although orographic GWs create strong GW hot spots in the stratosphere and lower mesosphere (e.g., Hendricks et al., 2014; X. Liu et al., 2019), they dissipate from dynamical instability or critical levels below the turbopause (e.g., Fritts et al., 2016). Thus, orographic GWs cannot propagate into the thermosphere and create the observed TAD hot spot. Vadas et al. (2019) analyzed TADs over the Southern Andes observed by the GOCE satellite on 5 July 2010. They found that nearly all of the GWs had such large horizontal phase speeds that they could not have come from below $z \sim 80$ km. Therefore, it is quite likely that the TAD hot spot observed over the Southern Andes/Antarctic Peninsula during wintertime results from the generation of secondary and tertiary GWs.

GW dynamics in the thermosphere is different from that in the middle atmosphere since here the predominant GWs are subject to strong molecular viscosity and heat conduction. As a result, only GWs having large intrinsic horizontal phase speeds and long vertical wavelengths can exist in the *F* region (Vadas, 2007). Hence, the usual assumptions of weak viscosity and the Boussinesq limit for the dispersion and polarization relations of GWs do not apply for GWs in the thermosphere. This limitation is overlooked in general circulation models (GCMs) when conventional GW parameterizations designed for the middle atmosphere are extended into the thermosphere. Furthermore, the conventional wisdom that the majority of relevant wintertime GWs above the mesopause has its sources in the troposphere is likely incorrect. Rather, the aforementioned multistep vertical coupling mechanism is the most plausible explanation for the wintertime GW observations in the thermosphere. However, conventional GW parameterizations exclude secondary and tertiary GWs by definition. Due to the implicit assumption of steady-state GW energy equations (see discussion in Becker, 2017), these parameterizations exclude intermittency and propagation of individual GW packets; furthermore, spatial localization is excluded due to the single-column approximation. On the other hand, intermittency and spatial localization are pivotal for secondary and tertiary GW generation by the body-force mechanism (Vadas et al., 2018). Therefore, to parameterize secondary and tertiary GWs, new frameworks for GW schemes need to be developed that can accommodate the necessary mechanisms. Such developments are currently in progress (e.g., Muraschko et al., 2015; Olbers et al., 2019).

To the best of our knowledge to date, numerical simulation of the multistep vertical coupling mechanism requires the explicit description of GWs from the surface up to high altitudes in the thermosphere. While idealized models with regional geometry resolve small-scale GWs and acoustic waves up to the thermosphere (e.g., Fritts et al., 2020; Heale et al., 2018, 2020), a high-resolution GCM (which includes radiative transfer, moisture cycle, orography and boundary layer processes, as well as synoptic to planetary-scale waves and realistic tidal variations) is needed to address the role of secondary and tertiary GWs in a global context. The primary purpose of this study is to present such a GCM and to provide some validation and first applications.

The following conceptual aspects should be taken into account when constructing a whole atmosphere GCM with the explicit simulation of GWs: (1) Since the mean flow effects of GWs are closely linked to viscosity and heat conduction, it is important to account for dissipative effects in a hydrodynamically consistent fashion (Becker, 2009). In GCMs that do not extend into the thermosphere, dissipative effects are due to explicit subgrid-scale (SGS) macroturbulent diffusion, giving rise to large eddy simulations (LES). When the thermosphere is resolved, the same model equations must allow for a mixture of LES and direct numerical simulations (DNS) at altitudes where the molecular viscosity exceeds the SGS diffusion coefficient. (2) Since GCMs are usually based on a dynamical core that solves fluid dynamical equations for a single

constituent, it is important to consistently extend the thermodynamic relations into the thermosphere where the resultant gas constant and heat capacities change significantly with altitude (e.g., H.-L. Liu et al., 2018). (3) Since even medium-scale GWs in the thermosphere deviate from the hydrostatic limit, it is important to take nonhydrostatic dynamics into account.

The GW-resolving GCM presented in this study extends from the surface to a pressure level of 6×10^{-9} hPa, corresponding to an altitude of ~ 450 km (depending on the thermospheric temperature). This model is a vertical extension of the former Kühlungsborn Mechanistic general Circulation Model (KMCM) and is abbreviated as HIAMCM (High Altitude Mechanistic general Circulation Model). The HIAMCM is a mechanistic model in the sense that the computation of radiation transfer and the tropospheric moisture cycle are simplified and that the model includes no thermospheric/ionospheric physical processes other than molecular diffusion and an ion drag parameterization. Except for water vapor and a simple proxy for tropospheric clouds, all minor constituents used in the radiation scheme are prescribed. In this study, the HIAMCM is used to simulate the neutral dynamics of the whole atmosphere for moderate solar maximum conditions (exospheric temperature of about 900–1000 K).

The remainder of this study is organized as follows. Section 2 and the appendices describe the model, including a validation of its global energetics. In section 3 we present and discuss model results regarding GWs in the middle and upper atmosphere. Section 4 gives a summary and some concluding comments.

2. Model Description

The HIAMCM is a free-running GCM based on a standard spectral dynamical core with a terrain-following vertical coordinate and a staggered vertical grid according to Simmons and Burridge (1981). We employ a triangular spectral truncation at total horizontal wavenumber 240, which corresponds to a horizontal grid spacing of ~ 55 km and a shortest resolved horizontal wavelength of $\lambda_h \sim 165$ km. The vertical level spacing is ~ 600 m between the boundary layer and 3×10^{-4} hPa ($z \sim 100$ km). The level spacing increases in the thermosphere and is ~ 5 km above ~ 300 km. Using 260 full model layers, the highest model layer is located at 6×10^{-9} hPa, corresponding to $z \sim 450$ km for temperatures of $T \sim 950$ K above ~ 250 km. We abbreviate this resolution as T240L260.

The HIAMCM includes simplified but nevertheless explicit representations of the relevant components of an atmospheric climate model: radiative transfer, water vapor transport, large-scale condensation and moist convection, the full surface energy budget including a slap ocean, macroturbulent and molecular horizontal and vertical diffusion, and ion drag. Further details of these parameterizations are given in the appendices. Table 1 summarizes the symbols that are introduced in this section.

Using a vertical hybrid coordinate, η , the pressure is defined as $p = a + b p_s$, where a and b are prescribed functions of η and p_s is the surface pressure. We have $p = p_s$ for $\eta = 1$ (hence, $a(1) = 0$ and $b(1) = 1$), and $p = 0$ for $\eta = 0$ (hence, $a(0) = 0$ and $b(0) = 0$). We set $\eta = p/p_{00}$ above $p \sim 90$ hPa (~ 15 km), where $p_{00} = 1,013$ hPa is the mean surface pressure at sea level. In the continuous notation, the momentum, enthalpy, and specific humidity equations solved by the HIAMCM can be written as

$$\begin{aligned} \partial_t \mathbf{v} &= \mathbf{v} \times (f + \xi) \mathbf{e}_z - \eta \partial_\eta \mathbf{v} - \nabla \left(\frac{\mathbf{v}^2}{2} + \Phi_s + \int_\eta^1 \frac{R T}{p} \partial_\eta p \, d\eta' + \Phi^{nh} \right) - \frac{R T}{p} \nabla p \\ &+ \frac{1}{\partial_\eta p} \nabla \cdot \left((K_h + \nu_h) \mathbf{S}_h + K_f \mathbf{S}_{hf} \right) + \frac{g}{\partial_\eta p} \partial_\eta \left(\frac{p}{R T} (K_z + \nu_z) \frac{g p \partial_\eta \mathbf{v}}{R T \partial_\eta p} \right) - \mathbf{D} \mathbf{v} \\ c_p \partial_t T &= -c_p (\mathbf{v} \cdot \nabla + \eta \partial_\eta) T + \frac{R T \dot{p}}{p} \\ &+ \frac{g}{\partial_\eta p} \partial_\eta (S - S_r + D - U) + Q_{large} + Q_{conv} \\ &+ \frac{c_p}{\partial_\eta p} \nabla \cdot \left(\partial_\eta p \left(\frac{K_h + \nu_h}{0.7} \nabla T + K_f \nabla T_f \right) \right) \\ &+ \frac{g c_p}{\partial_\eta p} \partial_\eta \left(\frac{p}{R T} \left(\left(\frac{K_z + \nu_z}{P_r} + \frac{\nu_z}{0.7} \right) \frac{g p \partial_\eta T}{R T \partial_\eta p} - \frac{K_z g}{P_r c_p} \right) \right) \\ &+ (K_h + \nu_h) (\mathbf{S}_h \nabla) \cdot \mathbf{v} + K_{hf} (\mathbf{S}_{hf} \nabla) \cdot \mathbf{v} + (K_z + \nu_z) \left(\frac{g p \partial_\eta \mathbf{v}}{R T \partial_\eta p} \right)^2 + \mathbf{v} \cdot (\mathbf{D} \mathbf{v}) \end{aligned} \quad (1)$$

$$(2)$$

Table 1

List of Symbols Introduced in Section 2 to Define the Model Equations

λ, ϕ, t	longitude, latitude, time
p, p_s	atmospheric pressure, surface pressure
T, T_s	atmospheric temperature, surface temperature
η	vertical coordinate; $\eta = 0$ for $p = 0$ and $\eta = 1$ for $p = p_s$
a, b	prescribed functions of η to compute $p = a + b p_s$
\mathbf{v}, ξ	horizontal wind, relative vorticity
f, g	Coriolis parameter, gravity acceleration
∇, \mathbf{e}_z	horizontal gradient operator, unit vector in vertical direction
Φ_s	orography (geopotential at the surface)
Φ^h, Φ^{nh}	hydrostatic geopotential, nonhydrostatic correction
R, c_p	gas constant, heat capacity per unit mass at constant pressure
K_h, K_z	macroturbulent horizontal and vertical kinematic viscosities
ν_h, ν_z	molecular horizontal and vertical kinematic viscosities
S_h	horizontal shear tensor of \mathbf{v}
\mathbf{v}_f, T_f	filtered horizontal wind and temperature
K_{hf}, S_{hf}	hyperdiffusion coefficient and horizontal shear tensor of \mathbf{v}_f
D, P_r	ion drag tensor, macroturbulent vertical Prandl number
S, S_r	downward and upward (reflected) solar energy flux densities
D, U	downward and upward infrared energy flux densities
q	specific humidity (mass mixing ratio of water vapor)
Q_{large}, Q_{conv}	latent heating due to large-scale and convective condensation
C_{large}, C_{conv}	corresponding condensation rates
\mathbf{v}_L, T_L, q_L	\mathbf{v}, T, q at the lowest full model layer
c_{surf}	heat capacity of the surface
c_D, c_{Dq}	surface exchange coefficients
f_{sens}, f_{latent}	surface fluxes of sensible and latent heat
ℓ	latent heat per unit mass
h_{ocean}	prescribed oceanic lateral heat flux convergence

$$\begin{aligned} \partial_t q = & -(\mathbf{v} \cdot \nabla + \dot{\eta} \partial_\eta) q - C_{large} - C_{conv} \\ & + \frac{1}{\partial_\eta p} \nabla \cdot (\partial_\eta p (K_h + \nu_h) \nabla q) + \frac{g}{\partial_\eta p} \partial_\eta \left(\frac{p}{R T} (K_z + \nu_z) \frac{g p}{R T} \frac{\partial_\eta q}{\partial_\eta p} \right) \end{aligned} \quad (3)$$

Here, \mathbf{v} is the horizontal wind vector, ∇ the horizontal gradient operator, f and ξ are the Coriolis parameter and relative vorticity, T is the temperature, and Φ_s the surface geopotential. Furthermore, $\dot{\eta}$ and \dot{p} are the material velocities of η and p , respectively. Partial derivatives with respect to time and η are denoted as ∂_t and ∂_η , respectively. A correction for the geopotential, Φ^{nh} , is included on the right-hand side (rhs) of Equation 1 to take nonhydrostatic effects into account. This nonhydrostatic geopotential is specified in Appendix B.

The macroturbulent horizontal and vertical diffusion coefficients are denoted as K_h and K_z , while ν_h and ν_z are the corresponding contributions from molecular kinematic viscosity (see Appendix A). We use an anisotropic kinematic molecular viscosity with $\nu_h > \nu_z$ above $z \sim 300$ km, which is further explained in Appendix A (see also Figure A2). $K_f(\eta)$ is a prescribed horizontal diffusion coefficient that is used for hyperdiffusion. The horizontal diffusion scheme in Equation 1 employs the horizontal shear tensor in spherical coordinates, S_h (Becker & Burkhardt, 2007), and a corresponding shear tensor, S_{hf} , that corresponds to a filtered horizontal wind, \mathbf{v}_f (Brune & Becker, 2013). We use $\partial_\eta p$ as a surrogate for density to compute horizontal diffusion in the hybrid vertical coordinate system. This is required in the troposphere to preserve angular momentum and energy on the model surfaces (Becker & Burkhardt, 2007). The last term on the rhs of Equation 1 is the ion drag (e.g., C. H. Liu & Yeh, 1969). In the current version of the HIAMCM we use

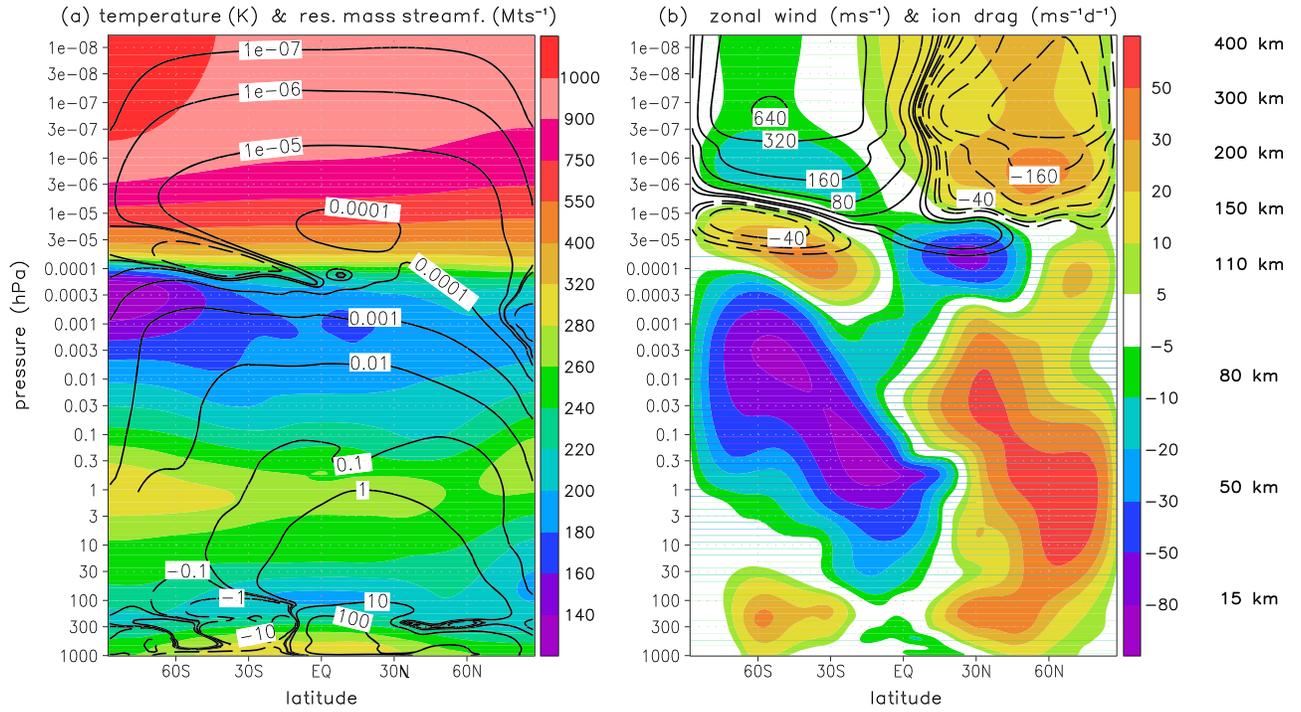


Figure 1. Simulated zonal-mean temperature (a) and zonal wind (b) during 1–30 January (colors). Contours in (a) show the residual mass stream function, which is drawn for $+10^{-7}$, $+10^{-6}$, $\pm 10^{-5}$, $+10^{-4}$, $+10^{-3}$, $+10^{-2}$ Mts^{-1} above 1 hPa, and for ± 0.1 , ± 1 , ± 10 , $+100$ Mt s^{-1} below 0.03 hPa. Contours in (b) show the zonal-mean zonal component of the ion drag which is drawn for ± 10 , ± 20 , ± 40 , ± 80 , ± 160 , ± 320 , $+640$ $\text{m s}^{-1} \text{day}^{-1}$. The vertical coordinate is the model's hybrid vertical coordinate times 1,013 hPa. Approximate geometric altitudes are given on the rhs of panel (b).

the idealized parameterization of Hong and Lindzen (1976) for solar maximum conditions (see Appendix E). No artificial sponge layer is included in the HIAMCM.

The dynamical terms in the sensible heat equation (first row of Equation 2) are as usual. Radiative heating (first term in the second row of Equation 2) includes the incoming (downgoing) and reflected (upgoing) solar energy flux densities, S and S_r , and the upgoing and downgoing longwave radiative energy flux densities, U and D . The last two terms in the second row of Equation 2 represent the latent heating rates due to large-scale condensation, Q_{large} , and moist convection, Q_{conv} . The radiation and convection schemes used in the HIAMCM are described in Appendices F and C, respectively. Horizontal and vertical diffusion of heat (third row of Equation 2) are specified analogously to the diffusion terms in the momentum equation (T_f is a horizontally filtered temperature). The turbulent vertical diffusion of heat is scaled with a turbulent Prandtl number, P_r , which is specified in Appendix A1. We use a Prandtl number of 0.7 for molecular diffusion (e.g., Vadas, 2007), as well as for macroturbulent horizontal diffusion.

The last row of Equation 2 consists of the turbulent shear production and molecular frictional heating rates due to horizontal and vertical momentum diffusion, as well as the frictional heating due to ion drag (Becker, 2003, 2009, 2017). Note that $K_h(\mathbf{S}_h \nabla) \cdot \mathbf{v} = K_h |\mathbf{S}_h|^2 \geq 0$, which is in accordance with the second law. The shear production from hyperdiffusion (second term in the last row of Equation 2) is not positive definite but is usually much smaller than $K_h |\mathbf{S}_h|^2$.

The budget for the mass mixing ratio of water vapor (or specific heat, Equation 3) includes horizontal and vertical advection, large-scale and convective condensation rates (C_{large} and C_{conv} , see Appendix C), and horizontal and vertical diffusion. We use the method of Schlutow et al. (2014) to compute the transport and diffusion of water vapor.

The hydrostatic formula in the hybrid vertical coordinate system is

$$\partial_\eta \Phi^h = -\frac{RT}{p} \partial_\eta p, \quad (4)$$

where $\Phi^h = g \bar{z}$, $g = 9.81 \text{ m s}^{-2}$, and \bar{z} is the geopotential height. For postprocessing of model data, the geopotential height is converted into the geometric height, z , using $z = \bar{z} (a_e + \bar{z})/a_e$, where a_e is the Earth radius. The model equations are completed by the continuity equation:

$$\partial_t \partial_\eta p + \nabla \cdot (\partial_\eta p \mathbf{v}) + \partial_\eta (\partial_\eta p \dot{\eta}) = 0, \quad (5)$$

the evolution equation for the surface pressure and expressions for the vertical velocities,

$$\partial_t p_s = - \int_0^1 \nabla \cdot (\partial_\eta p \mathbf{v}) d\eta \quad (6)$$

$$\dot{\eta} = - \frac{1}{\partial_\eta p} \left(b \partial_t p_s + \int_0^\eta \nabla \cdot (\partial_{\tilde{\eta}} p \mathbf{v}) d\tilde{\eta} \right) \quad (7)$$

$$\dot{p} = b (\mathbf{v} \cdot \nabla) p_s - \int_0^\eta \nabla \cdot (\partial_{\tilde{\eta}} p \mathbf{v}) d\tilde{\eta}, \quad (8)$$

by flux boundary conditions for the vertical diffusion terms in Equations 1–3:

$$\left[- \frac{p}{R T} (K_z + \nu_z) \frac{g p \partial_\eta \mathbf{v}}{R T \partial_\eta p} \right]_{surf} = \left[\frac{p}{R T} \right]_L C_D \mathbf{v}_L \quad (9)$$

$$\left[- \frac{p}{R T} \left(\left(\frac{K_z}{P_r} + \frac{\nu_z}{0.7} \right) \frac{g p \partial_\eta T}{R T \partial_\eta p} - \frac{K_z}{P_r} \frac{g}{c_p} \right) \right]_{surf} = \left[\frac{p}{R T} \right]_L C_D (T_L - T_s) = f_{sens} \quad (10)$$

$$\left[- \frac{p}{R T} (K_z + \nu_z) \frac{g p \partial_\eta q}{R T \partial_\eta p} \right]_{surf} = \left[\frac{p}{R T} \right]_L C_{Dq} (q_L - q_s) = f_{latent}, \quad (11)$$

and by the surface energy budget,

$$c_{surf} \partial_t T_s = [S - S_r + D - U]_{\eta=1} + c_p f_{sens} + \ell f_{latent} + h_{ocean}. \quad (12)$$

In Equations 9–11, the index “ L ” denotes the upper edge of the Prandtl layer, which is identified with the lowest full model layer. The latent heat per unit mass is $\ell = 2.5 \times 10^6 \text{ J/kg}$. The specific humidity at the surface, q_s , is defined in Equation C4 (see Appendix C). Furthermore, $c_{surf}(\lambda, \phi)$ and $h_{ocean}(\lambda, \phi)$ are the prescribed heat capacity of the surface and the lateral oceanic heat flux convergence, respectively (see also Becker & Vadas, 2018, their Figure 1). Details of the exchange coefficient at the surface are given in Appendix A. The flux boundary conditions (9)–(11) are completed by corresponding conditions at the model top where the fluxes are set to 0. Likewise, the vertical velocity (7) is subject to kinematic boundary conditions at the surface and at the model top (Simmons & Burridge, 1981).

Because of the change of atmospheric composition with increasing altitude, the gas constant, R , and the heat capacity at constant pressure, c_p , are no longer constant in the thermosphere. We account for the variations of R and c_p by assuming a monotonic dependence of the gas constant on the pressure. In this case, the heat capacity at constant pressure becomes a function of the temperature alone. This constraint is derived in Appendix D. Since the HIAMCM employs sensible heat as prognostic thermodynamic variable and pressure as a vertical coordinate above the tropopause, the functions $R(p)$ and $c_p(T)$ as given in Appendix D allow for a straightforward extension of the dynamical core into the thermosphere.

Semi-implicit time stepping is performed with regard to the actual global-mean temperature, and it incorporates linear contributions from the nonhydrostatic correction, as well as from horizontal and vertical diffusion of momentum, sensible heat, and water vapor using globally averaged diffusion coefficients (see Appendix A and Figure A1).

In the following we show first HIAMCM results for the northern hemispheric winter. We extracted an initial condition for 1 December from the free-running T240L220 KMCM used in Vadas and Becker (2019). We then added 40 model layers and used the spectral expansion coefficients of the uppermost layer from the

T240L220 KMCM to populate the additional layers. The HIAMCM T240L260 was then integrated for another nine model weeks. The same procedure was applied to the southern winter, using an initial condition from the former KMCM for 1 July. In each case, equilibrated dynamical states were reached after a few weeks of integration. Results presented in this paper refer to the integration periods from 1 to 30 January and from 1 to 20 August, which are the currently available model data with equilibrated dynamics.

3. Results

3.1. Zonal-Mean Picture

Before analyzing the simulated GWs in detail, we consider the usual zonal-mean diagnostics. The colors in Figure 1 show the zonal-mean temperature and zonal wind averaged in January. The model result for the lower and middle atmosphere is consistent with current knowledge based on either GW-resolving GCMs (e.g., Becker & Vadas, 2018; Watanabe & Miyahara, 2009) or comprehensive chemistry-climate models with parameterized GWs (e.g., Smith, 2012, Figures 1 and 2). That is, the cold summer mesopause and warm winter stratopause are linked to the residual circulation (contours in panel a). The first wind reversals in the summer MLT from westward to eastward flow (at $\sim 10^{-4}$ hPa in Figure 1a) and from eastward to westward flow (at $\sim 5 \times 10^{-6}$ hPa in Figure 1a) are too high in altitude by 1–2 scale heights each when compared to CIRA-86 (Fleming et al., 1990). This shortcoming is discussed below further. In contrast to models with parameterized GWs, a GW-resolving model does not simulate a reversal from eastward to westward flow with increasing height in the winter polar MLT on average, which is more consistent with observational data (see again Smith, 2012, Figures 1 and 2).

The simulated temperature strongly increases with height between about 100 and 150–200 km to an exospheric temperature of ~ 900 –1000 K. This is consistent with the global mean energy budget depicted in Figure F1b, showing an approximate balance between extreme ultraviolet (EUV) heating from the Sun and cooling due to downward molecular heat conduction above $z \sim 200$ km. The thermospheric mean zonal wind is westward in summer and eastward in winter. The latter is consistent with a mean temperature decrease from the summer to the winter pole, which is expected from a radiatively determined state. However, this temperature gradient is reduced by a summer-pole-to-winter-pole circulation at $z \sim 200$ –450 km that is largely driven by the zonal component of the mean ion drag (contours in Figure 1b). These thermospheric patterns are consistent with corresponding results from other global models (e.g., H.-L. Liu et al., 2018; Miyoshi et al., 2014). Due to the lack of ion chemistry, geomagnetic forcing, and energetic particle precipitation, the HIAMCM does not reproduce the observed westward mean winds in the thermosphere over the northern polar cap during wintertime (Drob et al., 2015).

Figure 2a shows the EPF divergence from all resolved waves (colors) along with the contribution from the resolved GWs (contours). This GW drag is identified as the EPF divergence due to horizontal wavenumbers smaller than $\sim 2,000$ km (horizontal wavenumbers larger than 20). The simulated GW drag in the middle atmosphere compares well with other model estimates (e.g., Fomichev et al., 2002; Smith, 2012). However, the eastward GW drag around the summer polar mesopause and the summer mesopause itself is too high in altitude by about 10 km. This is related to the wind reversal being too high in altitude as well, as mentioned above. The likely reasons for this shortcoming are (1) that the present model tuning is not sufficient, for example, regarding the turbulent diffusion scheme that induces GW dissipation, and (2) that the model resolution is still too coarse to resolve the GWs necessary for a more realistic eastward GW drag in the summer MLT. On the other hand, the model produces realistic dynamics of the winter middle atmosphere, including the westward GW drag due to orographic and nonorographic GWs from the upper stratosphere to the upper mesosphere poleward of about 30°N (Figure 2a). Furthermore, there is a significant eastward GW drag in the winter mesopause region and lower thermosphere. This feature was found previously for the southern winter middle atmosphere and was shown to be due to secondary GWs (Becker & Vadas, 2018).

In that work, the secondary GWs were generated by imbalances from the intermittent body forces (i.e., horizontal accelerations) resulting from the primary GWs that dissipate in the winter stratopause region. The mean wind then favored the upward propagation of eastward secondary GWs, which dissipate in the mesopause region and in the lower thermosphere. The current model simulation confirms that such an eastward GW drag is also found in the northern winter hemisphere. As discussed in Becker and Vadas (2018), this eastward GW drag from secondary GWs is the main reason why GCMs with resolved GWs usually do not

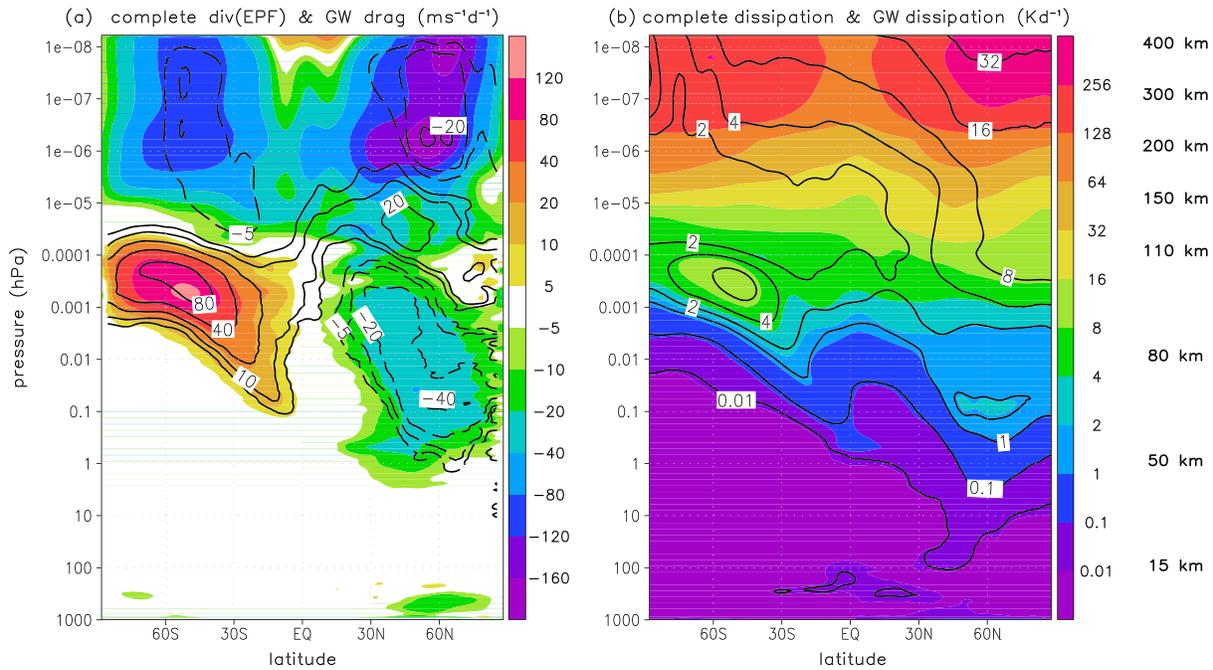


Figure 2. (a) Simulated wave driving during 1–30 January. Colors show the divergence of the EPF due to all resolved scales. The contours (± 5 , ± 10 , ± 20 , ± 40 , $+80 \text{ m s}^{-1} \text{ day}^{-1}$) show the EPF divergence due to horizontal wavelengths shorter than $\lambda_h \sim 2,000 \text{ km}$, which we interpret as the resolved GW drag. (b) Simulated dissipative heating during January. Colors show the frictional heating (from horizontal and vertical momentum diffusion, and from ion drag, see Equation 2) due to all scales. The contours (0.01 , 0.1 , 1 , 2 , 4 , 8 , 16 , and 32 K day^{-1}) show the contribution from horizontal wavelengths shorter than $\lambda_h \sim 2,000 \text{ km}$, which we interpret as the dissipation due to resolved GWs.

simulate an artificial reversal from eastward to westward mean flow over the winter polar cap in the mesopause region. Indeed, the mean zonal wind pattern in the winter MLT as simulated in our former GW-resolving model versions and in the HIAMCM compares well with other GW-resolving GCMs (H.-L. Liu, 2017; H.-L. Liu et al., 2014; Watanabe & Miyahara, 2009). At this point we do not know what role GWs generated from the polar vortex due to spontaneous emission (Sato et al., 2012; Sato & Yoshiki, 2008; Shibuya et al., 2017; Triplett et al., 2017) play in the eastward GW drag around the winter polar mesopause.

The GW drag in the winter hemisphere in Figure 2a changes sign again at about 150–200 km altitude and remains westward above $\sim 200 \text{ km}$. Therefore, we infer that the thermospheric GWs above $\sim 200 \text{ km}$ are likely tertiary GWs, analogous to the findings of Vadas and Becker (2019) for the southern winter hemisphere.

The zonal-mean thermospheric wave driving above $z \sim 150 \text{ km}$ is governed by large scales (total horizontal wavenumbers $n \leq 20$) and is predominantly westward. We extracted the tidal waves as in Becker (2017) and computed the EPF divergence for different zonal wavenumbers. The analysis showed that in the lower thermosphere up to about 150 km, the semidiurnal tidal component (zonal wavenumber 2), which is generated in the middle and lower atmosphere, gives the strongest contribution. At greater heights, the westward EPF divergence is mainly due to the diurnal tide (zonal wavenumber 1) that is generated in the thermosphere by the solar heating. Thus, we find that ion drag and the EPF divergence from large scales (mainly tides) are the main drivers of the residual circulation in the thermosphere, with values of about $\pm(100\text{--}600) \text{ m s}^{-1} \text{ day}^{-1}$. The resolved GWs, on the other hand, account for westward EPF divergence that typically amounts to only $10\text{--}20 \text{ m s}^{-1} \text{ day}^{-1}$ above $z \sim 150\text{--}200 \text{ km}$.

Figure 2b shows the mean frictional heating from all horizontal scales (colors) and from GWs (horizontal wavelengths shorter than $\sim 2,000 \text{ km}$, contours). The frictional heating from GWs is the main contributor below about 100 km and results from both vertical and horizontal turbulent momentum diffusion. The simulated heating pattern in the upper mesosphere is largely consistent with former estimates based on observations and models (e.g., Becker, 2009; Fomichev et al., 2002; Lübken, 1997). In the thermosphere, the frictional heating is controlled by the large scales, mainly tides, and amounts to $200\text{--}400 \text{ K day}^{-1}$ at z

~ 200–400 km. This is consistent with the fact that also the EPF divergence in the thermosphere is mainly due to large scales. Combining Figure 2b with the global-mean results given in Figure A3 suggests that the damping of the tides in the thermosphere is mainly due to ion drag and to a lesser extent due to molecular viscosity. Since the averaged frictional heating in the thermosphere is mainly due to the tides, the generation and dissipation of the diurnal tide is the major, Lorenz-type energy cycle above about 150 km.

Further inspection of the model results shows that the GWs are mainly dissipated by momentum diffusion at all heights and that dissipation by ion drag is unimportant for the GWs. The reason is that diffusion is highly scale selective such that GWs, which have much smaller scales than the tides, are more efficiently damped by diffusion rather than by ion drag. In other words, the damping rates resulting from viscosity are much larger than those resulting from ion drag for most thermospheric GWs. This holds particularly for GWs having intrinsic periods less than ~ 2 hr, even during daytime solar maximum conditions (see section 3.2. in Vadas, 2007 and pages 241 and 242 in Gossard & Hooke, 1975).

Our simulated zonal-mean effects from the GWs in the thermosphere (e.g., westward zonal GW drag of about $10\text{--}20\text{ m s}^{-1}\text{ day}^{-1}$ and energy deposition of about $10\text{--}30\text{ K day}^{-1}$) are at least 1 order of magnitude smaller compared to the previous model results of Yiğit et al. (2009). In that study, a conventional GW parameterization with lower stratospheric launch level parameters was extended into the thermosphere and showed very strong GW-mean flow interactions in the thermosphere. Miyoshi et al. (2014) used a GW-resolving GCM and also found strong zonal-mean GW drag of up to $+200\text{ m s}^{-1}\text{ day}^{-1}$ above 200 km. This drag is comparable in magnitude with the westward EPF divergence in Figure 2 from all waves (colors). However, this result of (Miyoshi et al., 2014, their Figure 3) does not contradict our result because those authors used zonal and meridional wavenumbers larger than 5 to define the GW drag, which includes horizontal wavelengths up to 8,000 km, whereas our definition of GWs includes only total horizontal wavenumbers larger than 20 (horizontal wavelengths shorter than 2,000 km). Thus, large-scale inertia GWs and tidal components are included in the specific definition of the GW drag used in Miyoshi et al. (2014). Furthermore, the expression for the complete EPF divergence (e.g., Equations (15) and (16) in Becker, 2017) must be used to estimate the wave-mean flow interaction, because tides and inertia GWs have large Stokes drifts, which yield an EPF divergence that can be opposite to the vertical momentum flux convergence. However, Miyoshi et al. (2014) calculated only the vertical momentum flux convergence. Other differences between the HIAMCM and the model of Miyoshi et al. (2014) are that the HIAMCM has significantly higher spatial resolution, as well as a physically consistent treatment of horizontal diffusion (as opposed to utilization of hyperdiffusion).

3.2. Resolved GW Dynamics

Even though the zonal-mean picture is very useful in the middle atmosphere, its relevance is questionable in the thermosphere. The reason is that the instantaneous wind variations due to the tides and GWs are larger than or at least comparable to the mean winds there. In the following we analyze the simulated instantaneous GW variations in the thermosphere in relation to the tides and to the GW sources at lower altitudes.

Figure 3 shows snapshots of GW-related relative density fluctuations from the lower stratosphere to the upper thermosphere in the Northern Hemisphere on 4 January. Here, we used the total horizontal wavenumbers larger than 30 ($n > 30$, horizontal wavelength shorter than ~ 1,350 km) to define the GW perturbations. In panel (a), the deformed polar vortex in the lower stratosphere at $z = 15\text{ km}$ is indicated by the white contours representing the horizontal stream function. This panel features several GW packets. For example, there are GW packets over Alaska (and Northwest Canada) and over the Rocky Mountains (and farther downstream). These are presumably orographically generated, because the GW phases at 15 km are roughly perpendicular to the directions of the large-scale wind (southeastward over Alaska and eastward over the Rocky Mountains), and these directions are roughly the same in the lower troposphere for these two cases (not shown). This interpretation is also consistent with the fact that the GW packet over and downstream of the Rocky Mountains is no longer visible at $z = 60\text{ km}$ (panel b) where the large-scale wind has turned from eastward to southward, as can be inferred from comparing the horizontal stream functions in panels (a) and (b). Hence, the orographic GWs must have encountered a critical level. The GW packet at $z = 15\text{ km}$ over Alaska is also no longer visible at $z = 60\text{ km}$. This is presumably because in this region the background wind has become very weak (as is indicated by the larger distances between stream function contours compared to panel a), thereby inducing wave breaking and subsequent damping by turbulent diffusion for these

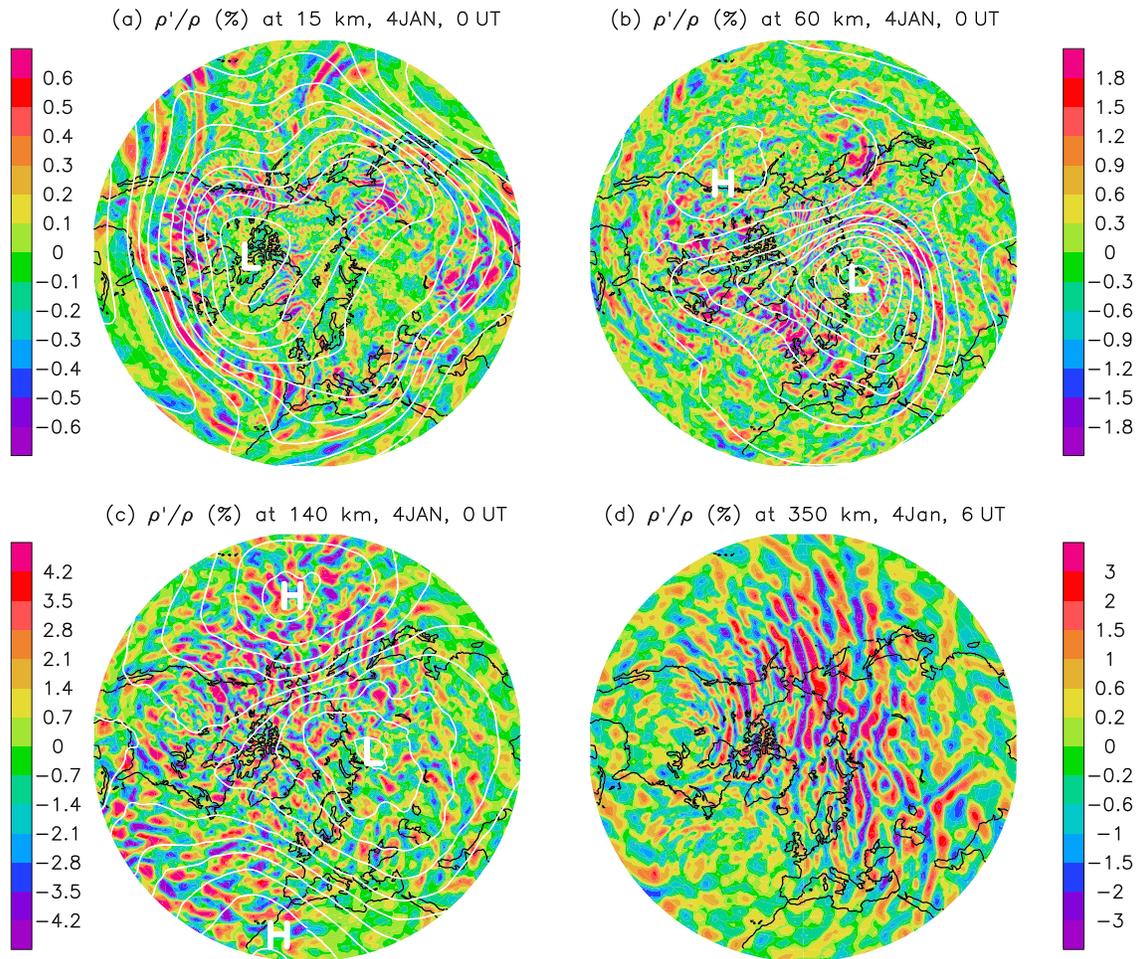


Figure 3. North polar projection (poleward of 20°N) of relative density perturbations (colors) from GWs with horizontal wavelengths shorter than $\lambda_h \sim 1,350$ km (total wavenumbers $n > 30$) at 15, 60, 140, and 300 km on 4 January: (a–c) 0 UT and (d) 6 UT. Here, the model data have been interpolated from the model surfaces to these geometric heights. The polar vortex at 15, 60, and 140 km is indicated by the horizontal stream function with white contours, whereby the minima and maxima are marked by gray capital letters L and H, respectively. In the Northern Hemisphere, the flow is clockwise (counterclockwise) around the maxima (minima) of the stream function.

orographic GWs. Note that in the HIAMCM, both the vertical and horizontal diffusion coefficients depend on the local and instantaneous Richardson number (see Equations A6–A11) such that the smallest-scale GWs that become dynamically unstable are strongly damped. This parameterizes the energy cascade at unresolved (subgrid) scales. For GWs having horizontal wavelengths larger than $\lambda_h \sim 400$ km, part of this cascade is resolved in the HIAMCM (Becker & Vadas, 2018).

In addition to the orographic GWs discussed above, Figures 3a and 3b feature inertia GWs that are likely associated with spontaneous emission. The GWs generated by that mechanism often show localized GW packets in the exit region of a jet with the GW phases approximately perpendicular to the streamlines (e.g., O’Sullivan & Dunkerton, 1995; Zhang, 2004), for example, at 15 km over the East Atlantic from North Africa to Brittany and at 15 km over the Pacific at about 170°E. Or they show spirals closely aligned with (approximately parallel to) the streamlines of the vortical flow (e.g., Sato & Yoshiki, 2008). This occurs, for example, in Figure 3a at $\sim 30^\circ\text{N}$ from the Pacific to the Atlantic Ocean over North America and in Figure 3b from the Eastern Mediterranean to Siberia.

From the lower thermosphere on, tidal variations have strong amplitudes while kinematic viscosity (molecular viscosity divided by density) increases rapidly with altitude. As a result, only GWs having large intrinsic phase speeds (of at least $\sim 150 \text{ m s}^{-1}$) can propagate to high altitudes in the thermosphere (Vadas, 2007). Model results show that wintertime primary and secondary GWs dissipate from dynamical instability in the

upper MLT (Becker & Vadas, 2018; Vadas & Becker, 2019). The resulting body forces lead to the generation of secondary and tertiary GWs that are visible as concentric rings of the GWs in horizontal snapshots above and below the body force (Vadas et al., 2003, 2018). Such an example is visible in Figure 3c (at $z = 140$ km) with the center of the concentric rings at about 90°W and 50°N (over the Rocky Mountains). At this altitude, the largest amplitudes are seen for the GWs propagating northwestward, westward, and southwestward away from this center. As indicated by the stream function contours in Figure 3c, the northwestward propagation direction is favored by the southeastward large-scale flow over Alaska and western Canada.

This picture is strongly variable in time due to the tidal motions. Moreover, the tidal winds have strong contributions from both the stream function and the velocity potential. Therefore, the stream function is no longer a valid representation of the background flow above $z \sim 150$ km. Figure 3d shows a snapshot at $z = 350$ km and 6 hr later than Figures 3a–3c in order to allow the GWs resulting from multistep vertical coupling to reach the upper thermosphere. Large-scale winds associated with the diurnal migrating tide (not shown) are strongly southward (northward) over North America (Siberia) at this time since it is LT midnight (noon) there (Roble & Ridley, 1994). The GWs resolved in the HIAMCM propagate approximately against this tidal wind, which is consistent with the observational findings of Crowley and Rodrigues (2012).

Figure 4 shows vertical snapshots that illustrate the GW propagation directions of some of the strong GW events seen in Figure 3 (using the same definition for GW perturbations). The longitude-height cross section in Figure 4a at 46°N focuses on the aforementioned lower stratospheric orographic GW event over the Rocky Mountains. The inclination of the GW phases in the stratosphere shows that the GW packet at 120 – 90°W and $z \leq 50$ km propagates westward with respect to the wind, thereby confirming that it is generated by eastward wind flow over the Rocky Mountains. The GW phases become flat above $z \sim 50$ km because the large-scale zonal wind turns westward in this regime (the black contour in Figure 4a marks $u = 0$). This induces the dissipation of orographic GWs beneath their critical level. According to Vadas et al. (2003) and Vadas et al. (2018), the resulting body force creates an imbalance in the background flow and thereby results in the generation of secondary GWs. These secondary GWs propagate upward and downward in all horizontal directions except perpendicular to the direction of the body force (i.e., they do not propagate northward or southward in this particular example). The amplitudes of the upward propagating secondary GWs are small at their generation altitude (e.g., $z \sim 50$ km here) but become large in the mesopause region. In the example depicted in Figure 4a, the secondary GWs are seen at $z \sim 60$ – 80 km and from about 120°W to 80°W . The increasing eastward wind from $z \sim 60$ to 100 km induces dissipation of the eastward propagating secondary GWs. The westward wind shear above $z \sim 100$ km in the longitude sector from 110°W to 80°W then induces the dissipation of the westward secondary GWs. As shown in Vadas and Becker (2019), the body forces that result from the dissipation of secondary GWs in turn lead to the generation of tertiary GWs. The tertiary GWs induced by the mountain wave event over the Rocky Mountains can be seen at altitudes above $z \sim 120$ km at 130 – 60°W in Figure 4a; these same GWs appear as concentric rings at $z = 140$ km in Figure 3c. Figure 4a furthermore indicates that some of the GWs in the broad tertiary GW spectrum dissipate by 140 km.

A closer inspection of the model data showed that the orographic event visible in Figure 4a over the Rocky Mountains started on 3 January at ~ 3 UT and that the secondary GWs first appeared in the upper mesosphere about 6 hr later (on 3 January at ~ 9 UT). Hence, in this particular example, the secondary GWs take about 6 hr to propagate from the upper stratosphere to the upper mesosphere, corresponding to a vertical group velocity of ~ 5 km/hr.

Figure 4b shows a latitude-height cross section at 65°E to further illustrate the aforementioned GW generation in the stratopause region by spontaneous emission. From Figure 3b, vortex-generated GWs propagate southward at $z \sim 60$ km, $\sim 65^\circ\text{E}$, and $\sim 35^\circ\text{N}$. (According to Sato and Yoshiki (2008), GWs generated from the polar night jet propagate equatorward on the equatorward side of the jet.) Therefore, the inclination of the GW phases in Figure 4b implies downward propagation at $z \sim 30$ – 60 km at 30 – 40°N . Furthermore, the meridional wind contours show a southward vertical shear at $z \sim 65$ – 80 km such that the upward and southward propagating GWs in this example dissipate by $z \sim 70$ – 80 km. The GW activity in the lower thermosphere in Figure 4b results from the lateral propagation of GWs generated by other sources but may also include secondary GWs created by body forces from the dissipation of the vortex generated GWs.

The latitude-height and longitude-height snapshots in Figures 4c–4f illustrate the meridional (c and d) and zonal (e and f) propagation of GWs from Figure 3d at approximately midnight, noon, 6 LT, and 18 LT,

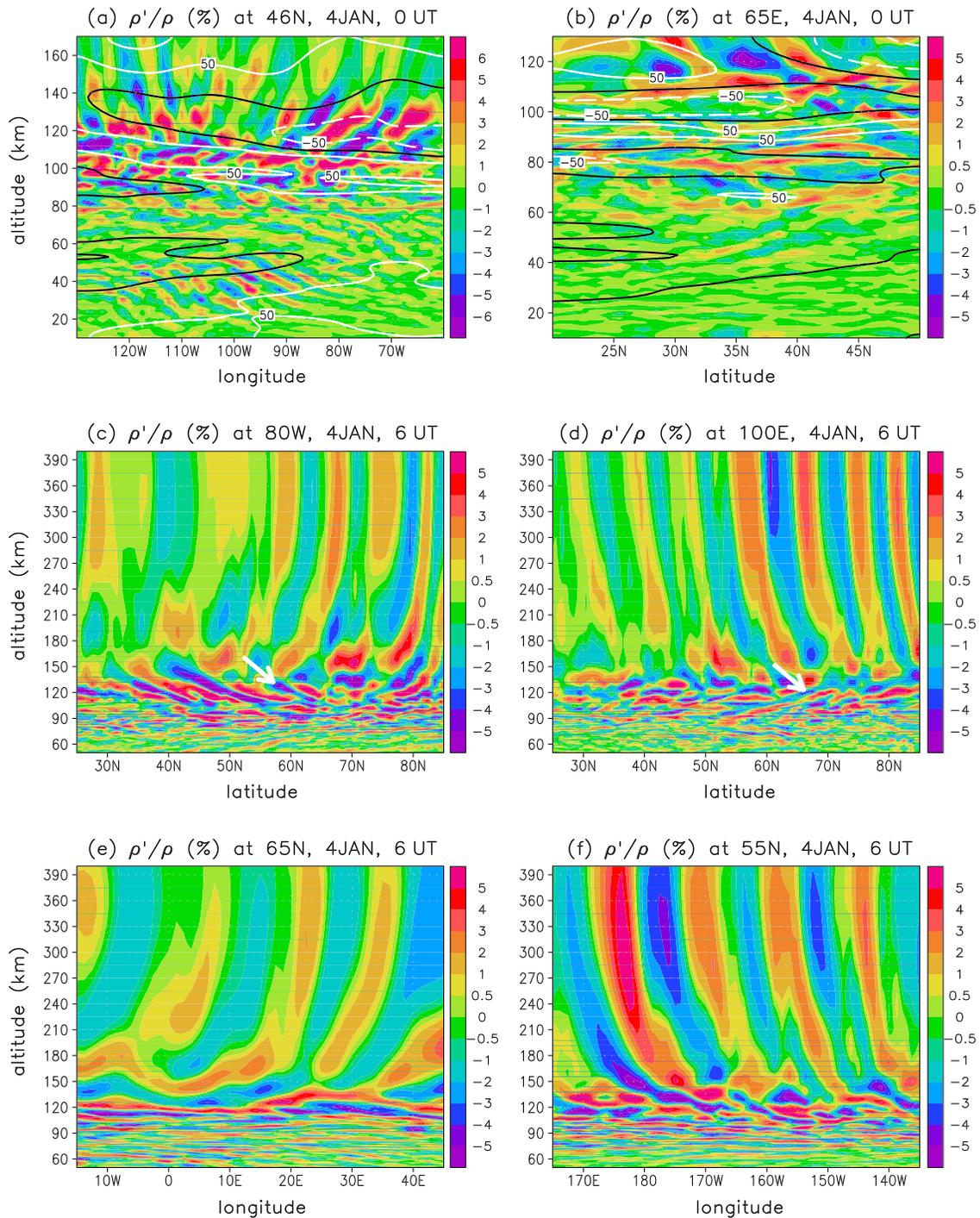


Figure 4. Horizontal-vertical cross sections of the GW density perturbations focusing on the GW events seen in Figure 3 on 4 January. Contours in panel a (panel b) show the large-scale zonal wind (meridional wind) for -50 , 0 , and $+50$ m s^{-1} using dashed white, solid black, and solid white contours, respectively. The white arrows in (c) and (d) mark fishbone structures. See text for more details.

respectively. The large-scale thermospheric wind (not shown) is primarily due to the diurnal tide and is approximately southward at midnight (c), northward at noon (d), westward at 6 LT (e), and eastward at 18 LT (f) in the Northern Hemisphere. The GWs at $z \geq 180$ km have very long vertical wavelengths, in agreement with linear viscous theory (Fritts & Vadas, 2008; Vadas, 2007). Importantly, the thermospheric GWs propagate approximately against the background wind, that is, northward, southward, eastward, and

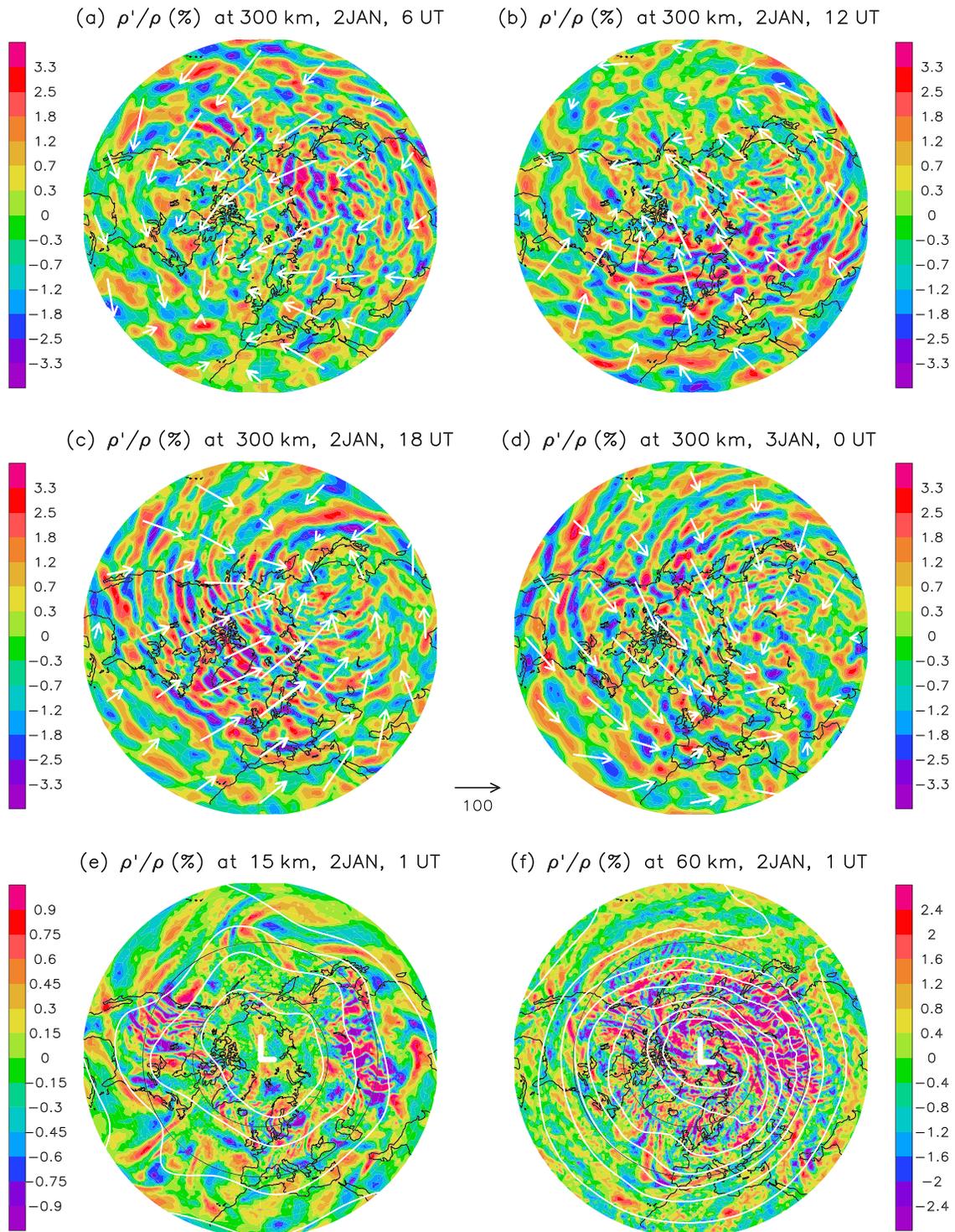


Figure 5. (a–d) North polar projection (poleward of 20°N) of 6-hourly consecutive snapshots of relative density perturbations (colors) owing to GWs having $\lambda_h < 1,350$ km at $z = 300$ km on 2–3 January. White arrows show the instantaneous large-scale horizontal wind field. A 100 m s^{-1} black arrow is shown between panels (c) and (d) as a scale. (e and f) Same as Figures 3a and 3b but on 2 January 1 UT and with larger color and contour intervals. The latitudes 40°N and 60°N are indicated by black circles.

westward in Figures 4c–4f, respectively, as seen by the inclination of the GW phases. In general, Figures 4c–4f indicate that larger GW scales dominate at $z \geq 150$ km compared to lower altitudes. This could be because tertiary GWs are generated at $z \sim 120$ – 160 km (see the fishbone structures at these altitudes, two of which are marked by white arrows in panels c and d) or because those GWs generated at lower altitudes having large horizontal wavelengths and phase speeds become dominant at $z > 150$ km.

The propagation of GWs in the thermosphere is further illustrated in Figures 5a–5d where we show 6-hourly snapshots of GW perturbations and the large-scale horizontal wind at $z = 300$ km on 2–3 January. We see a 24-hourly clockwise rotation of the predominant GW propagation directions that is linked to the clockwise rotation of the diurnal tide in the Northern Hemisphere. This HIAMCM result is typical for the Northern Hemisphere winter (not shown). Figures 5a–5d also show that the GW propagation is not precisely opposite to the tidal wind but lags somewhat (Crowley & Rodrigues, 2012).

Figures 5e and 5f illustrate the GWs and the polar vortex in the upper troposphere and around the stratopause at 2 January, 1 UT. At this time, the polar vortex was strong and therefore favored the vertical propagation of primary GWs into the stratopause region. In particular, a strong GW event is seen over Siberia/Mongolia at middle to high latitudes (panels e and f), which presumably gives rise to tertiary GWs in the thermosphere. These GWs are likely seen as concentric rings in Figure 5c, the center of which is located downstream over northeastern China. Additionally, large-amplitude GWs are also excited in the jet-exit region at $z = 15$ km over the Rocky Mountains (panel e) and are also seen in the stratopause region (panel f). There are some weak ring-like structures in Figure 5a at $z = 300$ km centered over the Rocky Mountains, which may indicate secondary or tertiary GWs induced by this event.

Given our arguments for the multistep vertical coupling mechanism to explain the medium- and large-scale GWs in the winter thermosphere at middle and high latitudes (Vadas & Becker, 2019), the occurrence rate and strength of these tertiary GWs should correlate with the strength of the polar vortex. Frissell et al. (2016) used a system of high-frequency radars in the North American sector to observe daytime medium-scale TIDs (MSTIDs) from 2012 to 2015. Such MSTIDs are presumably caused by GWs via neutral-ion collisions (e.g., Hocke & Schlegel, 1996; Nicolls et al., 2014; Vadas & Nicolls, 2009). Indeed, Frissell et al. (2016) found that the occurrence of the MSTIDs is strongest between the onset and the breakdown of the polar vortex and that wintertime MSTID activity is minimum when the polar vortex is weak.

A prerequisite for the HIAMCM to yield reliable results for the GWs in the thermosphere is that the observed tidal variations are reproduced. We tuned the solar heating in the thermosphere (see Appendix F) such that the tidal wind variations from the climatological wind model HWM14 (Drob et al., 2015) and the climatological temperature model MSIS (Hedin, 1991) are simulated reasonably well in the upper thermosphere. Figure 6 shows data from HWM14 and MSIS for 12UT on 5 January 2014, corresponding to solar maximum conditions with weak particle precipitation, and the corresponding HIAMCM data in early January at 59°N . While the overall structure and phases of the observation-based tidal variations are captured by the HIAMCM, there are some discrepancies, especially in the lower to middle thermosphere where little data was incorporated into HWM14 and MSIS. The differences in the middle to upper thermosphere are likely because HWM14 and MSIS are climatological models and HIAMCM incorporates variability from below.

We now determine the typical scales of the GWs simulated in the thermosphere by the HIAMCM. As an example we pick the equatorward propagating GWs over the North Atlantic on 2 January at 18 UT in Figure 5c. Such GWs would represent the aforementioned daytime MSTIDs observed by Frissell et al. (2016) over the North American sector. Figure 7 shows latitude-height (a and c) and latitude-time (b and d) cross sections at 45°W on 2 January. The upper panels show the propagation of GWs having horizontal wavelengths $\lambda_h = 400$ – $2,000$ km (horizontal wavenumbers 20–100), while the lower panels refer to $\lambda_h = 165$ – 400 km (horizontal wavenumbers 100–240). The inclination of the GW phases in panels (a) and (c) indicates very long vertical wavelengths and southward propagation in the thermosphere, as expected. The GWs having $\lambda_h = 400$ – $2,000$ km give the predominant contribution to the relative density fluctuations (see color scale at the rhs of each panel). In general, we find ground-based periods of $1 \text{ hr} \leq \tau \leq 2 \text{ hr}$, for horizontal wavelengths of $400 \text{ km} \leq \lambda_h \leq 2,000 \text{ km}$, and $20 \text{ min} \leq \tau \leq 40 \text{ min}$, for $165 \text{ km} \leq \lambda_h \leq 400 \text{ km}$.

In the keograms of Figures 7b and 7d, some of the GW phases are marked by black solid lines, and the corresponding GW packets are enumerated from 1 to 4. Assuming southward propagation, we obtain

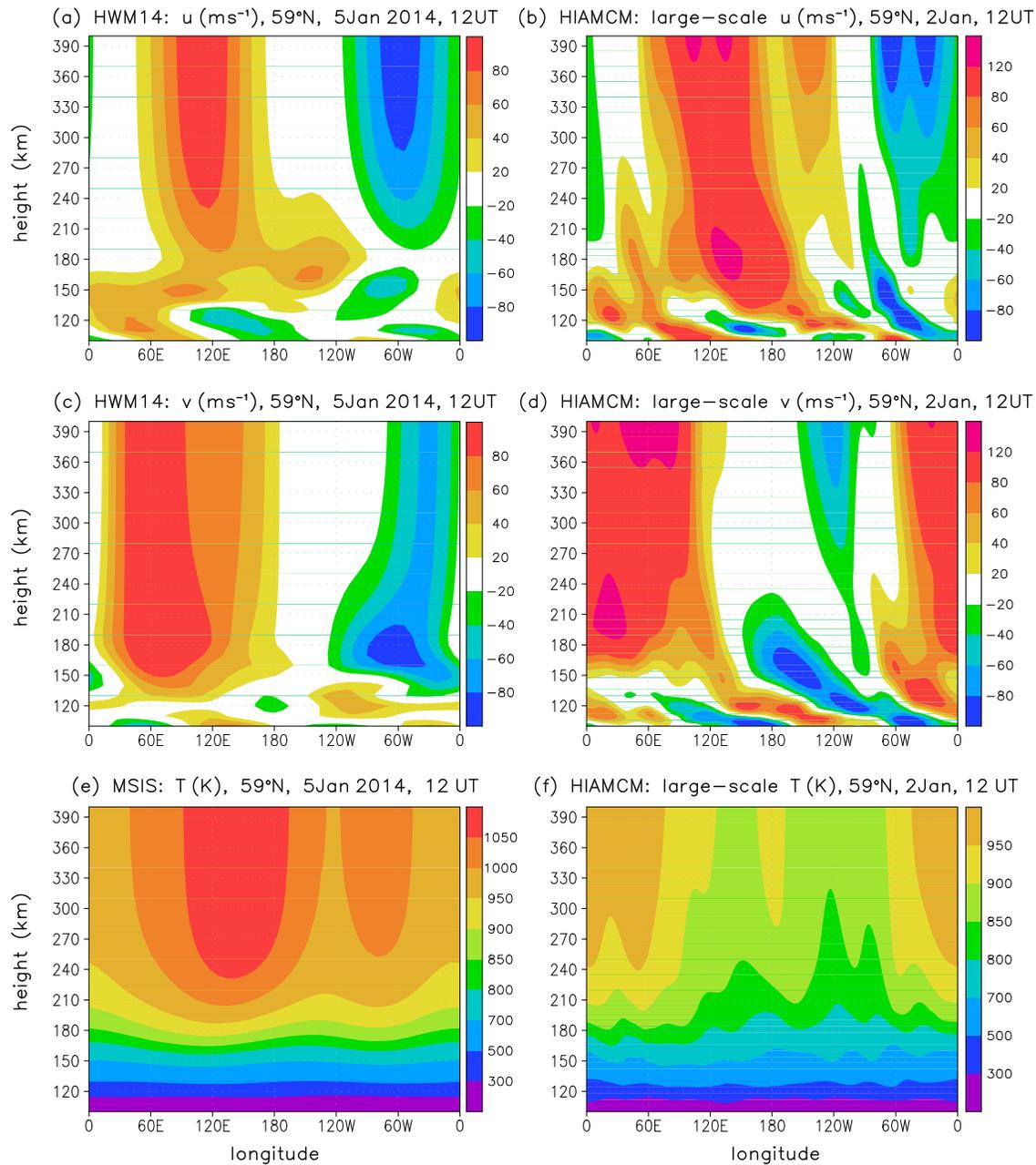


Figure 6. Comparison of tidal winds from the (a and c) HWM14 (Drob et al., 2015) and temperatures from the (e) MSIS (Hedin, 1991) with the (b, d, and f) HIAMCM simulation (large-scale flow including horizontal wavenumbers $n < 30$ only) for early January at 59°N .

horizontal wavelengths of $\sim 350\text{--}1,100$ km and ground-based periods of $\sim 25\text{--}90$ min. These correspond to ground-based phase speeds of $\sim 205\text{ m s}^{-1}$ for Packet 1, $\sim 245\text{ m s}^{-1}$ for Packet 2, and $\sim 195\text{ m s}^{-1}$ for Packets 3 and 4. These large phase speeds are comparable to the values given by Frisell et al. (2016) for MSTIDs. Furthermore, since these GWs propagate against the tidal background wind of $\sim 50\text{--}100\text{ m s}^{-1}$, their intrinsic phase speeds are $\sim 250\text{--}350\text{ m s}^{-1}$.

Are the GWs from Figure 7 compatible with the former concept that the predominant GWs in the thermosphere at middle and high latitudes are primary GWs generated in the troposphere (e.g., Yiğit & Medvedev, 2017)? To answer this question, we consider the dispersion relation for nonviscous GWs having medium or high frequencies:

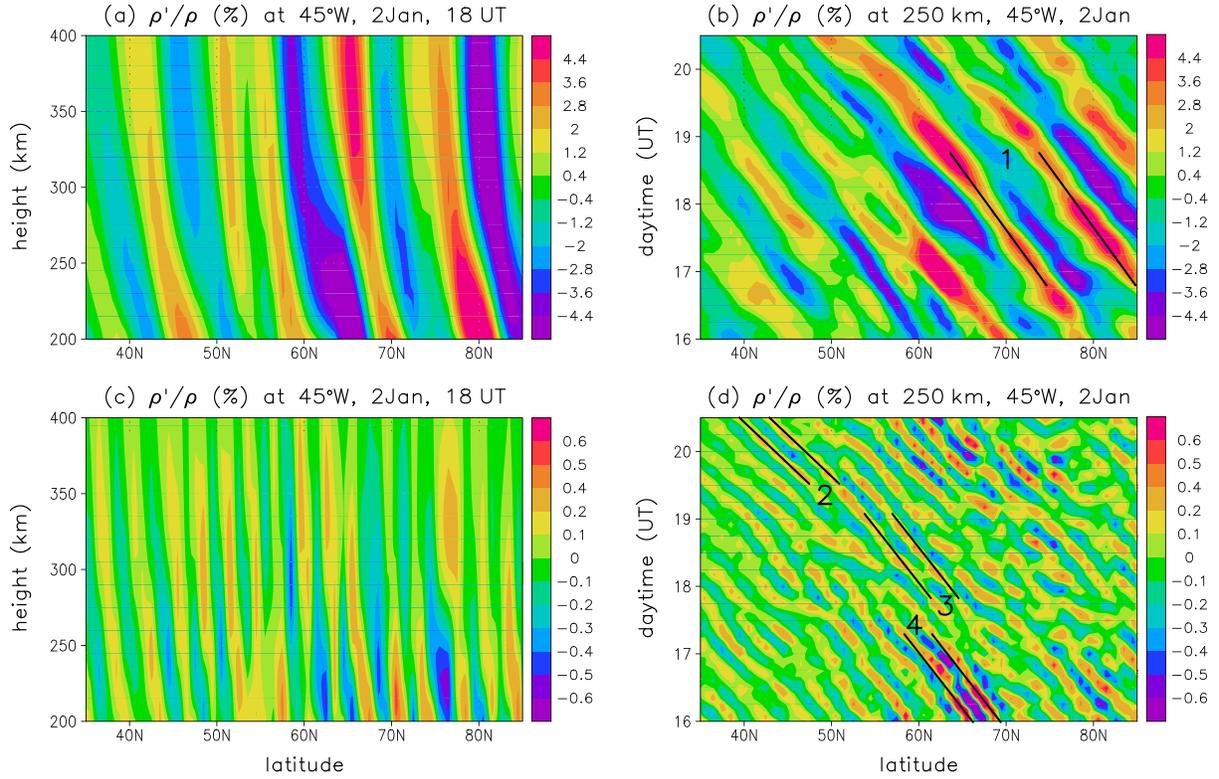


Figure 7. Southward propagating GWs in the thermosphere over the North Atlantic at 45°W around 18 UT (daytime). (a) Latitude-height cross section at 18 UT and (b) keogram at 250 km from 14 to 20.30 UT of the relative density fluctuations due to GWs having $400 \text{ km} < \lambda_h \leq 2,000 \text{ km}$. (c and d) Like (a) and (b) but for $165 \text{ km} \leq \lambda_h \leq 400 \text{ km}$. The black solid lines in panels (b) and (d) mark some GW packets that are enumerated from 1 to 4. Assuming southward propagation, we obtain the following approximate horizontal wavelengths (λ_h), ground-based periods (τ), and ground-based phase speeds ($c_h = \lambda_h/\tau$): $\lambda_h \sim 1,100 \text{ km}$, $\tau \sim 90 \text{ min}$, and $c_h \sim 205 \text{ m s}^{-1}$ for GW packet 1; $\lambda_h \sim 365 \text{ km}$, $\tau \sim 25 \text{ min}$, and $c_h \sim 245 \text{ m s}^{-1}$ for GW packet 2; and $\lambda_h \sim 350 \text{ km}$, $\tau \sim 30 \text{ min}$, and $c_h \sim 195 \text{ m s}^{-1}$ for GW packets 3 and 4.

$$\omega_I^2 = \frac{N^2 \mathbf{k}_h^2}{\mathbf{k}_h^2 + m^2 + 1/(4H^2)}. \quad (13)$$

Here, ω_I is the intrinsic frequency, N is the buoyancy frequency for constant background temperature, \mathbf{k}_h is horizontal wavenumber vector, m is the vertical wavenumber, and H is the density scale height. According to Vadas et al. (2019), a conservative estimate for the maximum possible value of the rhs of Equation 13 is obtained when both the horizontal and vertical wavelengths are large against $4\pi H$, yielding $\omega_I^2 < 4H^2 N^2 \mathbf{k}_h^2$. The reasoning is that $\mathbf{k}_h^2 \ll 1/(4H^2)$ for large enough λ_h and that λ_z becomes very large when a vertically propagating GW is close to reflecting. Using $N^2 = g^2/(c_p T)$ and $c_s^2 = c_p (c_p - R)^{-1} R T$ for the sound speed squared, we obtain the following constraint for the maximum intrinsic horizontal phase speed of a GW:

$$c_{I \text{ ph}}^2 < 4 \frac{(c_p - R)R}{c_p^2} c_s^2 = (r_c c_s)^2, \quad r_c = 2 \frac{\sqrt{(c_p - R)R}}{c_p}. \quad (14)$$

The black solid curve in Figure 8a shows the speed of sound from the HIAMCM during January and averaged from 50°N to 75°N. The dashed curve shows the maximum intrinsic phase speed according to Equation 14, where the factor r_c is given by the dashed curve in Figure 8b (including $R = R(p)$ and $c_p = c_p(T)$ from Appendix D). The blue curves show $v + 195 \text{ m s}^{-1}$ and $v + 245 \text{ m s}^{-1}$, where v is the large-scale meridional wind at 45°W, averaged from 50°N to 75°N and from 16:00 to 20:30 UT on 2 January (see Figure 7). Assuming that the (southward) ground-based horizontal phase speeds of the GWs are approximately constant with altitude, these blue curves are height profiles of the inferred intrinsic phase speeds of the GWs in Figure 7d. Because of the strong northward background winds at mesopause altitudes, the fastest GWs

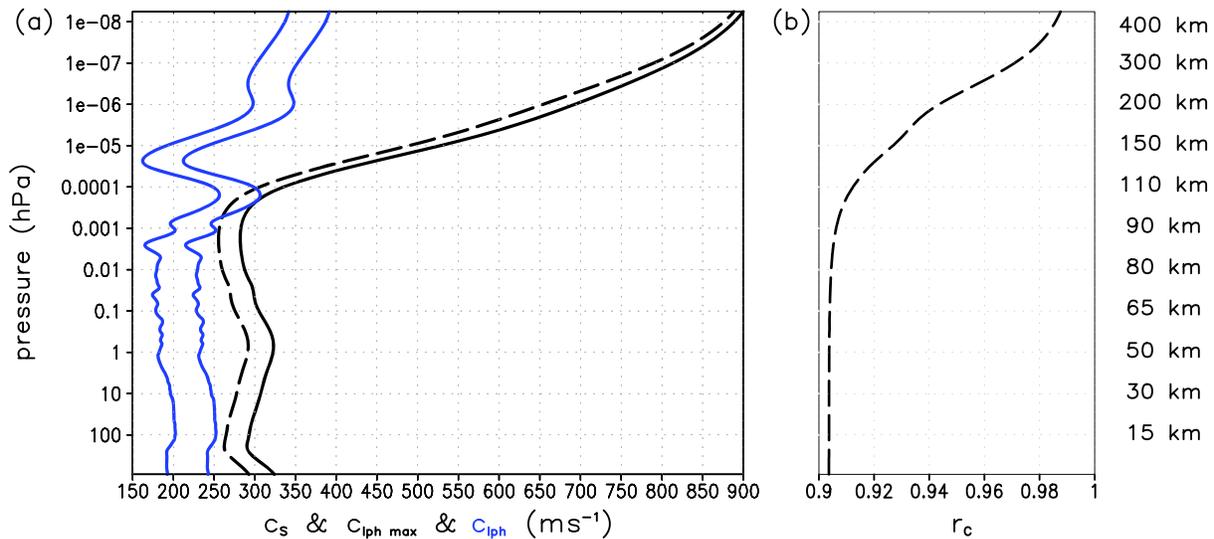


Figure 8. (a) Upper limit for intrinsic horizontal phases speeds of GWs. Solid black: speed of sound derived from the control simulation during January, averaged zonally and from 50°N to 75°N. Dashed black: maximum intrinsic phase speed according to Equation 14. Blue solid: $v + 195 \text{ m s}^{-1}$ and $v + 245 \text{ m s}^{-1}$, where v is the meridional wind at 45°W due to horizontal wavenumbers $n \leq 20$ at 45°W, averaged from 50°N to 75°N and from 16:00 to 20:30 UT on 2 January. (b) The factor r_c that converts the speed of sound into the maximum intrinsic phase speed (see Equation 14).

would have intrinsic horizontal phase speeds that exceed the speed of sound near the mesopause. Therefore, these GWs are most likely generated in the thermosphere. Moreover, even the slower GWs may meet the reflection criterion near the mesopause.

Figure 5c suggests that the concentric GWs analyzed in Figure 7 all originate from the same source over northeastern China, which creates ring-like structures of GWs propagating away from this source. Since we found that the faster GWs from that source are generated in the thermosphere, it is likely that all of the concentric GWs analyzed in Figure 7 are generated in the thermosphere. The same argument was also applied in Vadas and Becker (2019).

The situation depicted in Figures 5 and 7 is typical for the GWs simulated by the HIAMCM in the northern winter thermosphere at $z \sim 200\text{--}400$ km. Furthermore, the wintertime GWs at $z < 70$ km typically have horizontal phase speeds $< 100 \text{ m s}^{-1}$. Since these phase speeds are too slow for the corresponding GWs to survive viscous damping and propagate to $\sim 200\text{--}400$ km, this suggests that most of the wintertime GWs at these altitudes are likely generated above the mesopause.

3.3. Nonhydrostatic Effects

For the wave packets inspected in Figures 7c and 7d, the shortest intrinsic periods are less than 20 min. Since the buoyancy period is about 10–12 min in the 200–400 km region, these GWs are subject to nonhydrostatic effects. In the following we analyze the effect of the nonhydrostatic correction introduced in section 2 and Appendix B.

If high-frequency GWs are described using hydrostatic dynamics, the \mathbf{k}_h^2 term in the denominator of the dispersion relation (13) is neglected. This leads to an overestimation of the intrinsic frequency and, hence, of the ground-based frequency and phase speed. Furthermore, the sensible heat equation in the anelastic approximation yields (see Becker, 2017, Appendix)

$$T_a^2 = \frac{g^2}{c_p^2 \omega_f^2} w_a^2 \quad (15)$$

for the relation between the GW amplitudes of temperature (T_a) and vertical wind (w_a) in the conservative case. Assuming that a GW temperature amplitude is determined by energy conservation, hydrostatic dynamics will also overestimate the vertical wind amplitudes of high-frequency GWs. Hence, the HIAMCM with the nonhydrostatic correction should yield somewhat smaller horizontal phase speeds and smaller vertical wind amplitudes for GWs in the thermosphere as compared to the hydrostatic model version.

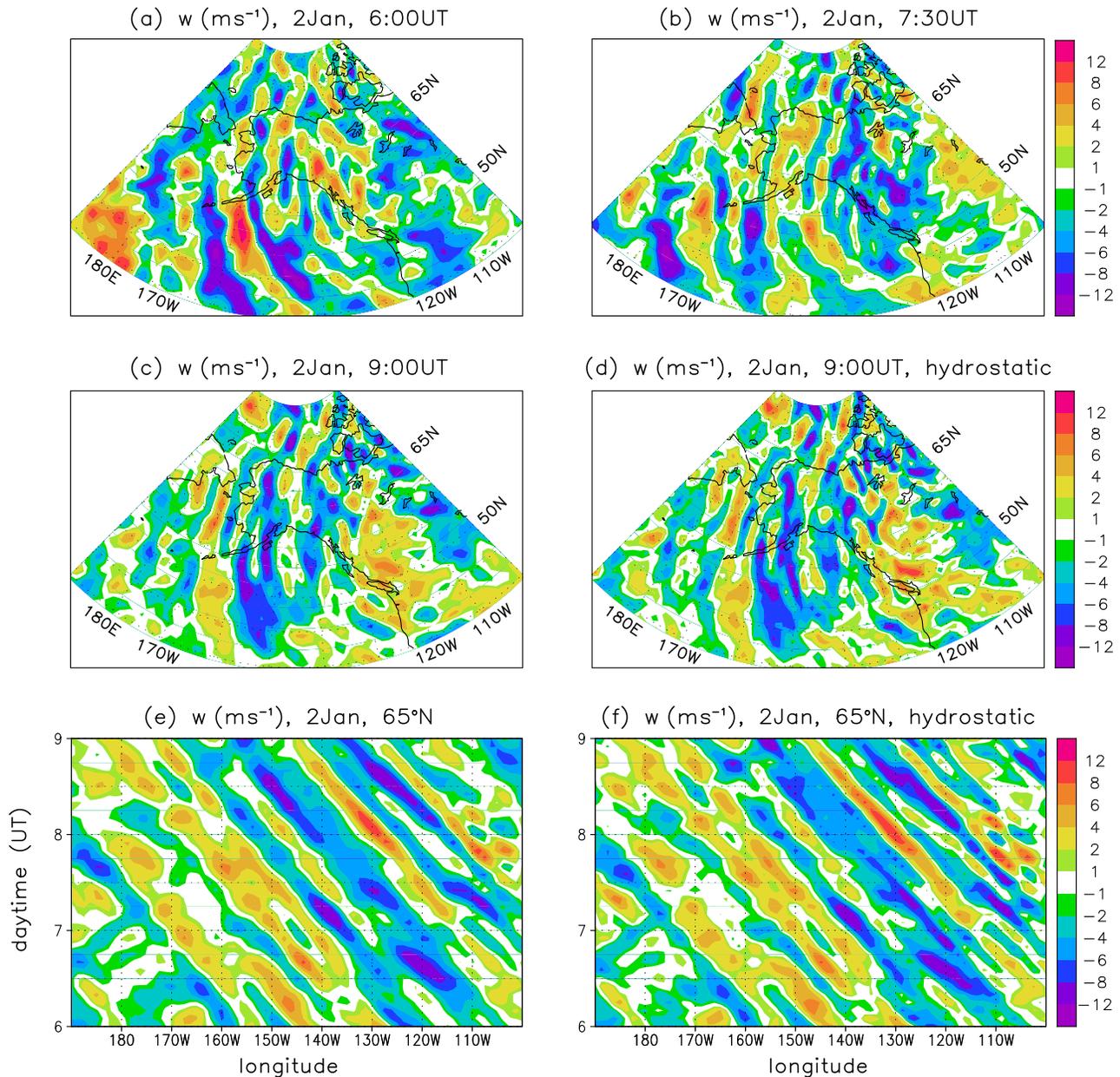


Figure 9. Vertical wind at $z = 320$ km over the eastern North Pacific at (a) 6, (b) 7:30, and (c) 9 UT on 2 January of the control simulation. (d) Same as (c) but for the hydrostatic model integration. (e) Keogram of the vertical wind at $z = 320$ km and 65°N from 6 to 9 UT. (f) Same as (e) but for the hydrostatic model integration.

We performed a sensitivity experiment to test these expectations. For this purpose we used 6 UT on 2 January as an initial condition to simulate the model for 3 hr without the nonhydrostatic correction. The corresponding time series is then compared to that of the control simulation (i.e., with the nonhydrostatic correction included). As an example we examine the westward propagating GWs visible in Figure 5a over the eastern North Pacific. Figures 9a–9c illustrate this westward propagation in terms of the vertical wind at $z = 320$ km at 6, 7:30, and 9 UT. The large-scale GW packet visible between 40°N and 55°N east of 170°W at 6 UT (panel a) has traveled about 15° to the west by 7:30 UT (panel b), corresponding to a phase speed of about 200 m s^{-1} . The predominant ground-based period of this packet is about 90 min. Taking into account that the background zonal wind is about $+100 \text{ m s}^{-1}$ (not shown, cf. to Figure 6), the predominant intrinsic period of this GW packet is about 1 hr, which is still in the hydrostatic regime. Accordingly, the comparison between the nonhydrostatic and the hydrostatic result at 9 UT in Figures 9c and 9d, respectively, does not indicate any

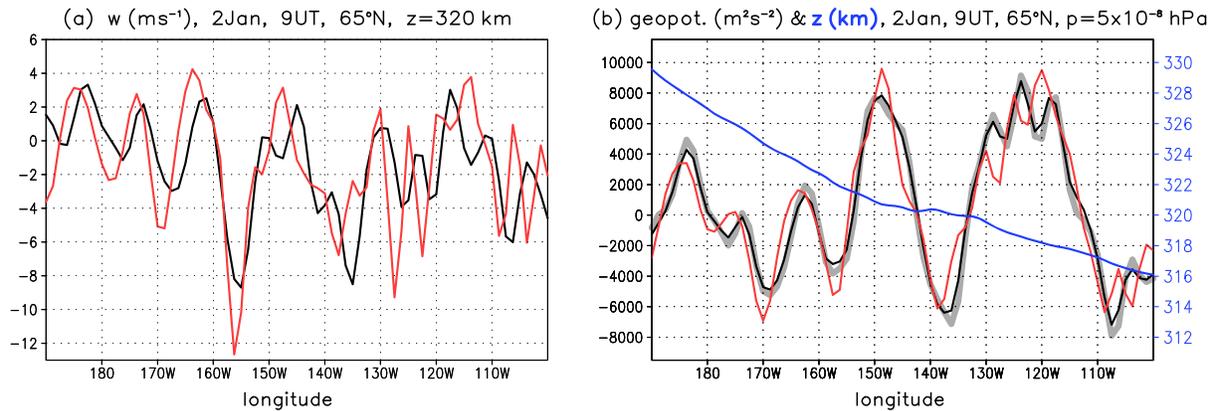


Figure 10. (a) Vertical wind at $z = 320$ km, 65°N , and 9 UT on 2 January from the control simulation (black curve) and from the hydrostatic model integration started at 6 UT (red curve). (b) Contributions to the geopotential at 5×10^{-8} hPa, 65°N , and 9 UT on 2 January: Hydrostatic geopotential due to GWs ($n > 20$) from the control run (gray) and the hydrostatic integration (red). The black curve gives the result when adding the nonhydrostatic correction from the control run to the gray curve. The large-scale geometric height is given by the blue curve (rhs axis).

notable differences for these large-scale GWs at 40 – 55°N . This is, however, different for the medium-scale GWs over Alaska and the Bering Sea at 55 – 65°N and $\sim 170^\circ\text{W}$. Here, the GWs from the hydrostatic simulation have propagated somewhat farther to the west compared to the nonhydrostatic case.

To further illustrate this difference, we show keograms of the vertical wind at $z = 320$ km and 65°N in Figures 9e and 9f. Assuming westward propagation, the typical ground-based phase speed of these medium-scale GWs is $\sim 230 \text{ m s}^{-1}$ and the corresponding intrinsic phase speed is $\sim 330 \text{ m s}^{-1}$. Since both simulations start with the same initial condition, differences become visible only after about 1 hr. At 9 UT, the GW phases have propagated farther to the west by several degrees when comparing the hydrostatic to the nonhydrostatic case. These differences can be better quantified from Figure 10a, which shows the vertical wind at $z = 320$ km, 65°N , and 9 UT from both simulations. The average phase difference is about 3° in longitude (corresponding to ~ 170 km) for the GW packet at 140 – 190°W . Furthermore, the hydrostatic simulation clearly shows $\sim 50\%$ larger vertical wind variations. These differences confirm that the hydrostatic case overestimates the GW phase speeds and vertical wind amplitudes.

The reason for these differences is further illustrated in Figure 10b, which shows different contributions to the geopotential at 5×10^{-8} hPa, 65°N , 9 UT. The blue curve in Figure 10b shows the large-scale geometric height (in km) that is due to total horizontal wavenumbers $n \leq 20$ (corresponding to $\lambda_h \geq 2,000$ km). This contribution is essentially identical in the two simulations; that is, after 3 hr of simulation from the same initial condition, the large scales in the two model integrations do not differ notably. The gray curve in Figure 10b is the hydrostatic part of the geopotential (Φ^h) in the control simulation that is due to horizontal wavenumbers $n > 20$ ($\lambda_h < 2,000$ km). Since the nonhydrostatic geopotential, Φ^{nh} (see Equation B4) consists mainly of small scales, we add Φ^{nh} to the gray curve, giving rise to the black curve. The differences between the gray and black curves indicate that the nonhydrostatic correction smooths the effective geopotential that forces the horizontal momentum (see Equation 1). This role of the nonhydrostatic correction is particularly evident from the red curve in Figure 10b, which shows the geopotential for $n > 20$ from the hydrostatic simulation. Compared to both the gray and the black curves, the red curve shows much more power at small scales, and its extrema are shifted to the west, reflecting again the unrealistically faster GW phase speed in the hydrostatic case.

3.4. GW Hot Spots in the Winter Thermosphere

The Southern Andes/Antarctic Peninsula region is a major hot spot for wintertime orographic GWs (e.g., Alexander & Teitelbaum, 2011; Eckermann & Preusse, 1999; Jiang et al., 2002; Wu & Waters, 1996). These waves are generated with large amplitudes when the predominantly eastward winds in the lower troposphere are strong. In addition, there is significant wintertime GW activity in stratosphere over the circumpolar Southern Ocean due to tropospheric GW generation by spontaneous emission (e.g., Ern et al., 2018; Hendricks et al., 2014). The wintertime orographic GWs usually break in the upper stratosphere and lower

mesosphere. Due to the strong intermittency of orographic GWs (Alexander et al., 2016), the Southern Andes/Antarctic Peninsula GW hot spot is also a region of strong generation of secondary GWs (Becker & Vadas, 2018; de Wit et al., 2017) and presumably also of tertiary GWs that result from the body forces due to the dissipation of secondary GWs (Vadas & Becker, 2019; Vadas et al., 2019). It is therefore likely that the corresponding GW hot spots observed by GOCE and CHAMP at $z \sim 250\text{--}450$ km (Forbes et al., 2016; Park et al., 2014; Trinh et al., 2018) result from tertiary GWs. Because GWs can propagate large horizontal distances in the thermosphere (e.g., Vadas & Liu, 2013), large GW activity far from the lower-altitude hot spot may result. This apparent discrepancy to the observed GW hot spots can be resolved by noting that the GW spectrum generated by a body force is broad (Vadas et al., 2003, 2018); it includes small-medium-scale GWs with high frequencies that propagate close to the zenith and GWs with smaller frequencies and larger scales that propagate closer to the horizontal. This suggests that the small-medium-scale GWs would form a GW hot spot close to the Southern Andes/Antarctic Peninsula in the upper thermosphere, while GWs having larger scales would travel larger horizontal distances and would therefore result in enhanced GW activity farther away. Only the small-medium-scale GWs having large-enough frequencies and vertical group velocities would be visible at $z \sim 250\text{--}450$ km near the geographical location of their sources. Hence, we expect an orographic GW hot spot in the stratosphere to be “mirrored” in the upper thermosphere through the multistep vertical coupling mechanism when we only examine the small- to medium-scale GWs.

Using the spectral decomposition of the temperature in terms of spherical harmonics, we compute the temperature variances due to different horizontal spectral regimes of the total horizontal wavenumber: background flow ($n = 0\text{--}20$ or $\lambda_h > 2,000$ km), large-scale GWs ($n = 21\text{--}50$ or $2,000 \text{ km} > \lambda_h \geq 800$ km), medium-scale GWs ($n = 51\text{--}120$ or $800 \text{ km} > \lambda_h \geq 330$ km), and small-medium-scale GWs ($n = 121\text{--}240$ or $330 \text{ km} > \lambda_h \geq 165$ km). Maps were produced by averaging the squared temperature fluctuations over horizontal areas of about $1,000 \times 1,000 \text{ km}^2$ and averaging in time.

Figure 11 shows the resulting temperature variance maps in the Southern Hemisphere during August at 50, 70, 90, and 110 km height. In the stratopause region (first row), the major orographic hot spots (Southern Andes/Antarctic Peninsula, New Zealand, Tasmania) show up most clearly in the medium-scale and small-medium-scale spectral regimes. Nevertheless, significant amplitudes are generated by the Southern Andes and Antarctic Peninsula even in the large-scale regime. With increasing altitude in the mesosphere, however, the temperature variance in the large-scale regime (left column) becomes roughly zonally symmetric. This means that these large-scale GWs are generated presumably by the dissipation of primary GWs resulting from flow over orography and spontaneous emission in the upper troposphere. Therefore, the sources of these mesospheric secondary GWs reside not only over the orographic hot spot regions but also over the circumpolar Southern Ocean (see also Becker & Vadas, 2018; Vadas & Becker, 2018). Moreover, these waves have small-enough vertical group velocities so that they propagate large horizontal distances while propagating from the lower to the upper mesosphere, resulting in a roughly zonally symmetric distribution (Figures 11g and 11j). Also note that this distribution of large-scale secondary GWs around the mesopause maximizes in the polar region. The likely reason is that propagation directions toward the south and the east are favored because the poleward mean meridional and eastward mean zonal winds decrease with altitude in the southern winter mesosphere. It is also likely that some GW activity seen in the large-scale regime above $z = 50$ km is caused by spontaneous emission from the stratospheric polar vortex (see Figures 3b and 4b).

The medium-scale and small-medium-scale temperature variances in Figure 11 (second and third columns) show maximum values somewhat downstream (eastward and poleward) of the Antarctic Peninsula and the Southern Andes at 70 and 90 km, which is similar to the orographic GW events analyzed in Becker and Vadas (2018) and Vadas and Becker (2019), indicating that these GWs are mainly secondary GWs. At $z = 110$ km, the Southern Andes/Antarctic Peninsula GW hot spot is only visible in the small-medium-scale spectral regime (Figure 11i). Using temperature data from the SABER (Sounding of the Atmosphere using Broadband Emission Radiometry) instrument onboard the TIMED (Thermosphere Ionosphere Mesosphere Energetics Dynamics) satellite, Trinh et al. (2018) and X. Liu et al. (2019) found that this hot spot is visible in the stratosphere and lower mesosphere, but not near the mesopause (e.g., at $z \sim 85\text{--}90$ km in Figure 2 of Trinh et al., 2018). A possible explanation is that the hot spot is partially smoothed out around the mesopause because of horizontal propagation of (secondary) GWs.

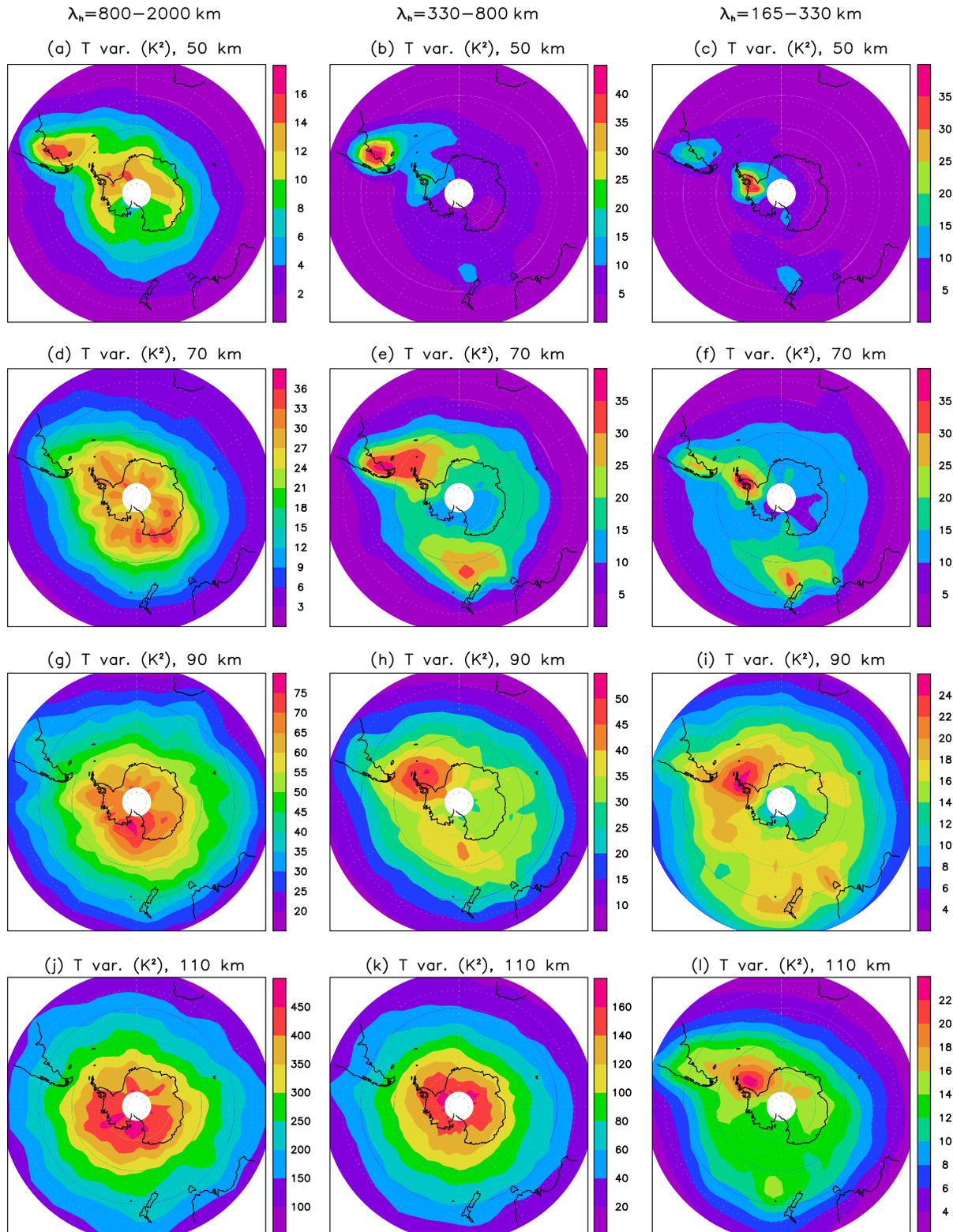


Figure 11. Simulated GW hot spots in the Southern Hemisphere (1–20 August, polar projection southward of 30°S). (a–c) Simulated temperature variances at 50 km from large-scale GWs ($\lambda_h = 800\text{--}2,000$ km), medium-scale GWs ($\lambda_h = 330\text{--}800$ km), and small-medium-scale GWs ($\lambda_h = 165\text{--}330$ km). (d–f, g–i, and j–l) Same as (a)–(c) but at 70, 90, and 110 km, respectively. Note the different color scales in the individual panels.

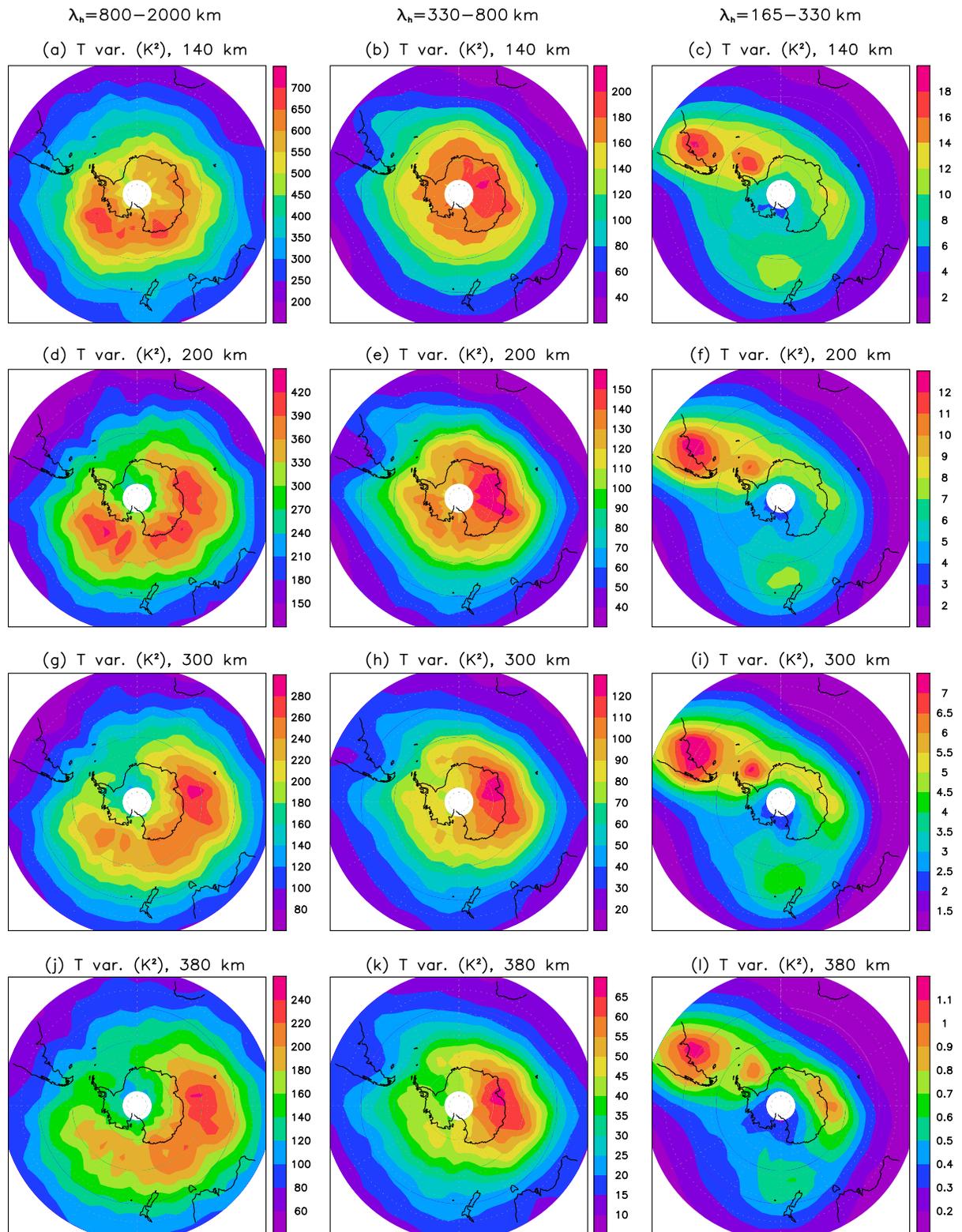


Figure 12. Same as Figure 11 but at 140, 200, 300, and 380 km.

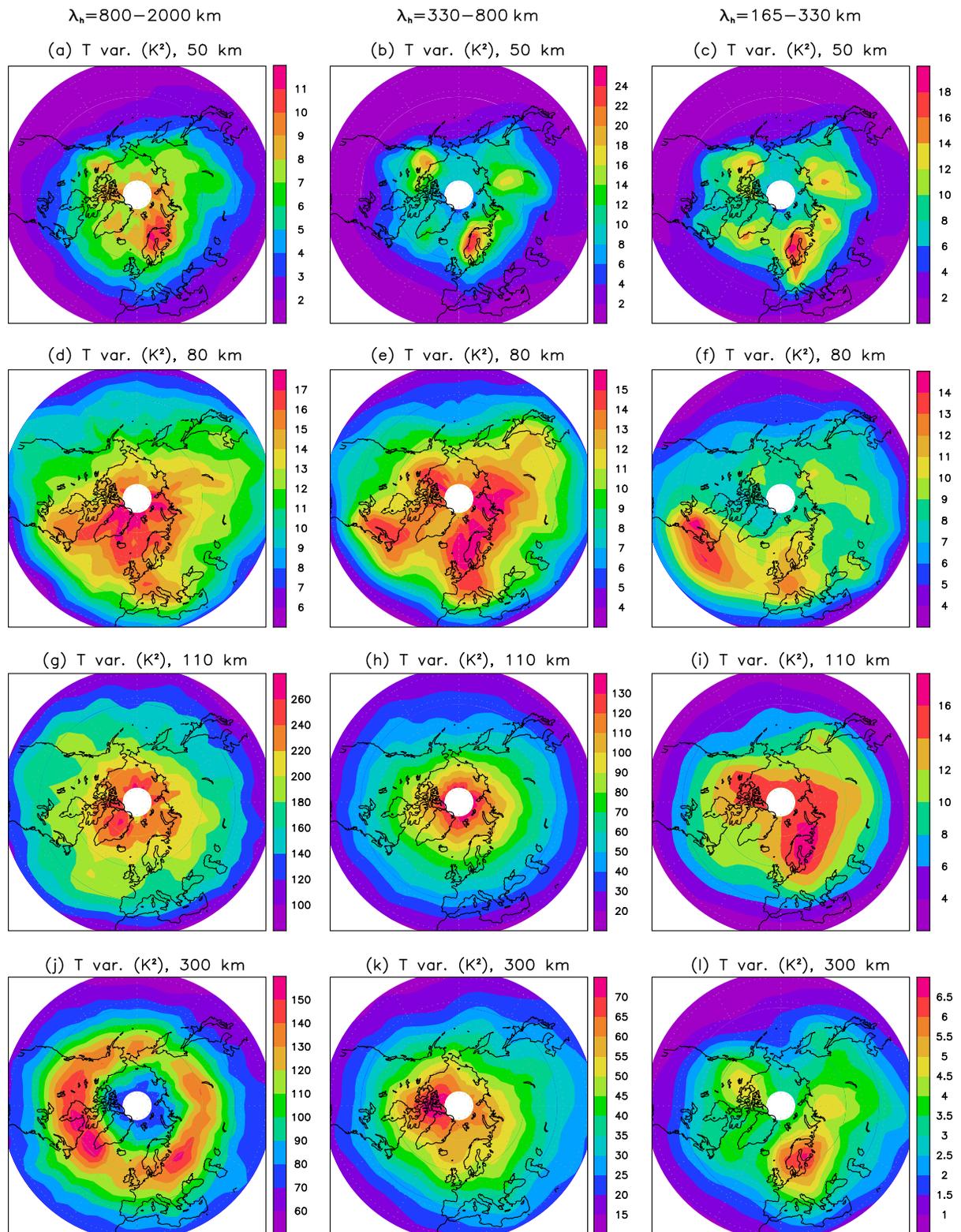


Figure 13. Same as Figure 11 but for the Northern Hemisphere winter (1–30 January, polar projection northward of 30°N) and at altitudes of 50, 80, 110, and 300 km.

The strongest and most intermittent body forces in the lower thermosphere are expected to result from the medium-scale and small-medium-scale GWs. This is because large-scale GWs having smaller intrinsic frequencies will generate weaker and less localized body forces due to their Stokes drifts and larger horizontal scales (e.g., Fritts & Alexander, 2003; Vadas & Becker, 2018). Hence, the generation of tertiary GWs is expected to be maximum somewhat eastward and poleward from the Southern Andes/Antarctic Peninsula GW hot spot on average, as is seen in Figure 1 of Vadas and Becker (2019). Furthermore, tertiary GWs having small-medium scales and large vertical group velocities should create a GW hot spot in the upper thermosphere.

Figure 12 shows the temperature variances analogous to Figure 11, but for the thermosphere. Indeed, the GW hot spot near the Southern Andes/Antarctic Peninsula is visible at all altitudes for medium-small-scale GWs (right column in Figure 12). This hot spot is strongest somewhat east of the Southern Andes in the thermosphere, whereas the corresponding maximum at $z = 90$ and 110 km (Figures 11i and 11j) is located east of the Antarctic Peninsula. The large-scale and medium-scale temperature variances in Figure 12 (first and second columns) are strongly indicative of GW propagation around and over the pole, creating a broad maximum at middle to high latitudes from about 30°E to 180°E that is most pronounced at $z \sim 300$ km. Such an additional broad maximum is also visible in the thermospheric results of Trinh et al. (2018, their Figures 2 and 6). The lower two rows in Figure 12 show that the small-medium-scale GWs are strongly damped from 300 to 380 km altitude, whereas the damping is much less for GWs with larger horizontal scales. This is because the complete damping of GWs due to molecular viscosity, which is roughly proportional to $\mathbf{k}_h^2 + m^2 - 1/(4H^2)$ (Vadas & Liu, 2013), is included in the HIAMCM. (GCMs usually do not include molecular viscosity in the horizontal diffusion scheme and therefore neglect \mathbf{k}_h^2 from molecular viscosity in the GW damping rate.) This strong viscous damping above $z \sim 300$ km of small-medium-scale GWs having typical intrinsic horizontal phase speeds of ~ 350 m s⁻¹ (see Figure 8) is consistent with results of Vadas (2007).

Figure 13 shows the model results of the GW hot spots at different altitudes in the Northern Hemisphere during January. Large-scale GWs in the stratopause region (panel a) are of minor relative importance compared to the Southern Hemisphere because the GW generation by spontaneous emission in the upper troposphere is weaker in the northern winter than in the southern winter due to weaker synoptic baroclinic Rossby waves at the expense of larger planetary Rossby wave activity. The medium-scale and small-medium-scale GW variances at 50 km (panels b and c) reflect the orographic GW hot spots where most of the GWs that propagate to 50 km are generated during the simulation period: Alaska, Greenland, Scandinavia, Ural, and Mongolia/Siberia. Due to the larger number of orographic hot spots, the sources of secondary and tertiary GWs are less localized in the northern than in the southern winter middle and upper atmosphere. Also note that the geographical distribution of GW activity in the northern winter middle atmosphere is subject to strong seasonal variability, as is most evident during sudden stratospheric warming events (e.g., Ern et al., 2016). Nevertheless, in our 30 day average of January model data, the variances due to small-medium-scale GWs at 300 km still reflect most of the orographic hot spots (Figure 13). Furthermore, the GWs in the southern winter thermosphere at 300 km (third row in Figure 12) show stronger maximum activity in all three spectral regimes than the GWs at 300 km in the northern winter thermosphere (last row in Figure 13). This is likely because the polar vortex is stronger in the southern winter than in the northern winter, thus causing the primary GWs to propagate to higher altitudes before dissipating and thereby creating stronger secondary and tertiary GW activity in the southern winter.

4. Summary and Conclusions

We described a new GW-resolving GCM called HIAMCM (High Altitude Mechanistic general Circulation Model) that extends from the surface (including orography, a slab ocean model, and the full surface energy budget) to the upper thermosphere (model top at about 450 km altitude). The model is mechanistic due to idealized computations of radiative transfer and moist convection, as well as due to the neglect of photo and ion chemistry. On the other hand, the model produces fairly realistic large-scale dynamics in the lower and middle atmosphere. In addition, GWs are simulated explicitly down to horizontal wavelengths of about 165 km. The following measures were relevant to achieve this goal: (1) applying a spectral dynamical core (which allows computation of all horizontal derivatives using the spectral transform method) with sufficient spatial resolution, (2) using the nonlinear Smagorinsky scheme for macroturbulent horizontal and vertical

diffusion such as to exploit the given numerical resolution and to simulate GW mean flow interaction in a self-consistent fashion (Becker, 2009), (3) completing horizontal and vertical diffusion by the molecular diffusion in the thermosphere, (4) extending the thermodynamic relationships consistently in the thermosphere, (5) completing the dynamical core by a nonhydrostatic correction, and (6) including global mean diffusion coefficients, the actual global mean temperature profile, and the nonhydrostatic correction in the semi-implicit time-stepping method. In contrast to other GW-resolving GCMs, the HIAMCM does not require an artificial sponge layer because the resolved GWs are mainly dissipated by macroturbulent diffusion up to the lower thermosphere and by molecular diffusion at higher altitudes. The HIAMCM fills the gap of explicitly simulating GWs in the thermosphere on a global scale, which is currently not feasible using comprehensive whole atmosphere GCMs (H.-L. Liu, 2017; H.-L. Liu et al., 2018). Note, however, that due to the limited spatial resolution of a GCM, the resolved GWs are spectrally biased toward medium scales, at least in the lower and middle atmosphere where turbulent diffusion is the dissipation mechanism.

Results from the HIAMCM support earlier findings that the GW activity in the winter thermosphere is predominantly due to secondary and tertiary GWs (Vadas et al., 2019). The secondary GWs are generated due to the imbalances (accelerations) that result when the primary GWs dissipate and thereby create localized body forces. Orographic GWs are efficient in this respect due to their large intermittency and localization in space. Tertiary GWs are generated in the upper MLT where the secondary GWs dissipate. In general, the GWs that propagate to the upper thermosphere have very long vertical wavelengths, as predicted from linear theory (Vadas, 2007). The intrinsic periods and horizontal phase speeds are larger than the corresponding ground-based quantities because the waves that survive at any given altitude mainly propagate against the mean background wind due to the in situ generated diurnal tide. This result is in agreement with observations of TIDs during quiet time conditions (Crowley & Rodrigues, 2012; Crowley et al., 1987; Friswell et al., 2016).

We find that in the winter thermosphere, the resolved GWs have ground-based periods of $1 \text{ hr} \leq \tau \leq 2 \text{ hr}$ for horizontal wavelengths of $400 \text{ km} \leq \lambda_h \leq 2,000 \text{ km}$ and $20 \text{ min} \leq \tau \leq 40 \text{ min}$ for $165 \text{ km} \leq \lambda_h \leq 400 \text{ km}$. The ground-based horizontal phase speeds are typically $200\text{--}250 \text{ m s}^{-1}$, and the intrinsic horizontal phase speeds in the considered cases are $250\text{--}350 \text{ m s}^{-1}$. Below the turbopause, a GW can have a maximum intrinsic horizontal phase speed of $\sim 90\%$ of the sound speed (Vadas & Crowley, 2010; Vadas et al., 2019). Many of the simulated GWs in the northern winter thermosphere would exceed this threshold near the mesopause. Thus, these GWs and the other slower GWs from the same source are most likely tertiary GWs generated above the mesopause.

The intrinsic periods of the resolved small-medium-scale GWs in the midthermosphere are larger than the buoyancy period by about a factor of about 1.3–3. Therefore, the hydrostatic approximation is no longer adequate to simulate these waves. Linear GW theory suggests that hydrostatic dynamics overestimates the frequencies, phase speeds, and vertical wind amplitudes of high-frequency GWs. In order to successfully simulate these GWs, we included a nonhydrostatic correction in the HIAMCM. A sensitivity simulation showed that this nonhydrostatic correction works successfully and leads to the expected behavior when compared to corresponding hydrostatic results.

Our model results show that GWs in the northern winter thermosphere having medium to large-scale horizontal wavelengths of order of $\lambda_h \sim 400\text{--}2,000 \text{ km}$ propagate thousands of kilometers horizontally while propagating from the lower to upper thermosphere. As a result, during the high-latitude winter, medium-to large-scale tertiary GWs propagate around and over the pole as they propagate to higher altitudes. At F region altitudes, many of these GWs propagate northward around midnight and southward around noon due to the clockwise rotation of the diurnal tide in the Northern Hemisphere. Note that southward propagating large-scale (LS) TIDs (LSTIDs) induced by such GWs could be misinterpreted as LSTIDs created from geomagnetic activity.

The small-medium-scale tertiary GWs tend to have small intrinsic periods and large vertical group velocities and therefore do not propagate very far horizontally from the source region. For the southern winter simulation, these GWs create a large hot spot in the middle and upper thermosphere over and somewhat downstream of the Southern Andes/Antarctic Peninsula. For our northern winter simulation, weaker stratospheric orographic hot spots occur over Alaska, Greenland, Scandinavia, Ural, and Mongolia/Siberia, and the small-medium-scale tertiary GW hot spots at $z \sim 300 \text{ km}$ are less structured and

weaker than in the southern winter. All these findings are in qualitative agreement with GOCE and CHAMP observations of quiet time GWs in the thermosphere. In the future, we will apply analysis methods to determine the sources of the GWs in the thermosphere such as was done in (Vadas & Becker, 2018, 2019).

Application of the zonal mean diagnostics in the Transformed Eulerian Mean (TEM) framework showed that even though the resolved GWs in the winter thermosphere have quite large amplitudes (10–20 m s⁻¹ in the vertical wind), the EPF divergence due to these waves is not very important. Tides and ion drag give the strongest forcing in the TEM zonal momentum equation. This is because the momentum deposition from the GWs that propagate against the background tidal winds largely averages out in the TEM picture. We also did not see a relevant EPF divergence due to resolved GWs in the summer thermosphere. This could, however, be different if high-frequency primary GWs generated by moist convection were taken into account in the HIAMCM, which might also help achieve a more realistic simulation of the summer polar mesopause.

Model results confirmed that the thermospheric GWs dissipate mainly from parameterized macroturbulent diffusion and, above ~200 km, mainly from molecular viscosity and heat conduction (Vadas, 2007; Vadas & Nicolls, 2009), whereas the tides dissipate mainly from ion drag. Since the dissipation of the tides is the major contribution to the dissipative heating in the thermosphere, it is the generation and dissipation of the diurnal tide that represents the major, Lorenz-type energy cycle in the thermosphere.

In the future we will further improve the HIAMCM by applying a higher resolution and a new SGS parameterization based on the dynamical Smagorinsky model (Schaefer-Rolffs & Becker, 2018), as well as a more sophisticated ion drag and further parameterizations of ionospheric processes. We are also currently developing the ability to nudge the large-scale dynamics in the lower atmosphere to reanalysis and/or forecast data so as to simulate GW events generated by flow over orography, as well as by jets and fronts, and their multistep coupling into the thermosphere for particular observational periods.

Appendix A: Macroturbulent and Molecular Diffusion

The HIAMCM with a T240L260 resolution does not include GW parameterization. Instead, GWs are simulated explicitly, subject to some biasing of the resolved GW activity toward larger scales than in the real atmosphere. The GWs are generated in the model by various processes: flow over orography (Becker & Vadas, 2018; Watanabe et al., 2006), nonlinear dynamics of synoptic-scale vortical flow (Plougonven & Zhang, 2014), and condensational heating (H.-L. Liu, 2017). GWs need to dissipate eventually in order to generate mean flow effects (EPF divergence, energy deposition, and mixing) in steady state. While dissipation in the thermosphere can be directly due to molecular viscosity and heat conduction, GWs break and generate turbulence with a forward energy cascade in the lower and middle atmosphere, as well as in the lower thermosphere. This insight is reflected, for example, by the saturation theory of Lindzen (1981), which is still the basis of many GW parameterizations used in community climate models. To formulate a high-resolution GCM with resolved GWs, we must parameterize the turbulent cascade and dissipation via a macroturbulent diffusion scheme.

We employ the classical Smagorinsky model (Becker & Burkhardt, 2007; Smagorinsky, 1993). This scheme is complemented by a weak hyperdiffusion to avoid a buildup of the energy spectra at the truncation wavenumber (Brune & Becker, 2013). Since the HIAMCM extends to high altitudes in the thermosphere, molecular viscosity and heat conduction are fully included. The framework of the diffusion scheme is given in section 2. In the following we describe details of the diffusion coefficients.

The horizontal shear tensor and its Frobenius norm are as follows:

$$\begin{aligned} \mathbf{S}_h &= ((\nabla + \mathbf{e}_z/a_e) \circ \mathbf{v}) + ((\nabla + \mathbf{e}_z/a_e) \circ \mathbf{v})^T - \mathbf{E} \vartheta \\ &= (\vartheta - 2 a_e^{-1} \partial_\phi v) (\mathbf{e}_\lambda \circ \mathbf{e}_\lambda - \mathbf{e}_\phi \circ \mathbf{e}_\phi) + (\xi + 2 a_e^{-1} \partial_\phi u) (\mathbf{e}_\lambda \circ \mathbf{e}_\phi + \mathbf{e}_\phi \circ \mathbf{e}_\lambda) \end{aligned} \quad (\text{A1})$$

$$|\mathbf{S}_h|^2 = (\vartheta - 2 a_e^{-1} \partial_\phi v)^2 + (\xi + 2 a_e^{-1} \partial_\phi u)^2. \quad (\text{A2})$$

Here, \mathbf{e}_z is the unit vector in the vertical direction, a_e the earth radius, and \mathbf{E} the unit tensor. Furthermore, λ and ϕ are longitude and latitude, and \mathbf{e}_λ and \mathbf{e}_ϕ are the corresponding unit vectors. The horizontal wind vector is decomposed as usual, $\mathbf{v} = u \mathbf{e}_\lambda + v \mathbf{e}_\phi$. Furthermore, $\xi = \mathbf{e}_z \cdot (\nabla \times \mathbf{v})$ and $\vartheta = \nabla \cdot \mathbf{v}$ are the relative vorticity and horizontal divergence in spherical geometry. The superscript ^T denotes the transpose.

Since the HIAMCM is a spectral model, we expand the horizontal wind vector into a series of spherical harmonics using

$$\mathbf{v} = - \sum_{n=1}^N \frac{a_e^2}{n(n+1)} \sum_{m=-n}^{+n} (\xi_{nm} (\mathbf{e}_z \times \nabla Y_{nm}) + \vartheta_{nm} \nabla Y_{nm}). \quad (\text{A3})$$

Here, the spherical harmonics are abbreviated as Y_{nm} , ξ_{nm} and ϑ_{nm} are the spectral expansion coefficients of ξ and ϑ , n is the total horizontal wavenumber, and m and $n - m$ are the corresponding zonal and meridional wavenumbers. We apply triangular spectral truncation at $N = 240$. The spectral expansion allows to define a filtered horizontal wind and the corresponding shear tensor used for additional hyperdiffusion as

$$\mathbf{S}_{hf} = ((\nabla + \mathbf{e}_z/a_e) \circ \mathbf{v}_f) + ((\nabla + \mathbf{e}_z/a_e) \circ \mathbf{v}_f)^T - \mathbf{E} \nabla \cdot \mathbf{v}_f \quad (\text{A4})$$

$$\mathbf{v}_f = - \sum_{n=1}^N f_n \frac{a_e^2}{n(n+1)} \sum_{m=-n}^{+n} (\xi_{nm} (\mathbf{e}_z \times \nabla Y_{nm}) + \vartheta_{nm} \nabla Y_{nm}), \quad (\text{A5})$$

where f_n is a filter in spectral space with $f_n = 0$ for $n \leq 200$ and increasing smoothly for higher wavenumbers such that $f_{240} = 1$. The explicit form of \mathbf{S}_{hf} is analogous to Equation A1.

The nonlinear Smagorinsky scheme specifies the horizontal and vertical diffusion coefficients based on the mixing-length concept of Ludwig Prandtl. Using the symbols l_h and l_z for the horizontal and vertical mixing lengths, we write the macroturbulent diffusion coefficients as

$$K_h = l_h^2 (|\mathbf{S}_h|^2 + S_{hmin}^2)^{1/2} (1 + 9 F(R_i - R_{i0})) \quad (\text{A6})$$

$$K_z = l_z^2 ((\partial_z \mathbf{v})^2 + S_{zmin}^2)^{1/2} F(\tilde{R}_i - R_{i0}), \quad \partial_z \mathbf{v} = \frac{-g p}{R T} \partial_\eta \mathbf{v} \quad (\text{A7})$$

$$F(R_i) = \begin{cases} \sqrt{1 - 18 R_i} & \text{for } R_i \leq 0 \\ 1/(1 + 9 R_i) & \text{for } R_i > 0. \end{cases} \quad (\text{A8})$$

Here, $S_{hmin}^2 = 4 \times 10^{-11} \text{ s}^{-2}$ and $S_{zmin}^2 = 3 \times 10^{-9} \text{ s}^{-2}$ are minimum squared horizontal and vertical wind shears ensuring that the spatial derivatives of the diffusion coefficients are always defined. The Richardson numbers in Equations A6 and A7 are

$$R_i = N^2 (\partial_z \mathbf{v})^{-2} = \left(\frac{T}{c_p} - \frac{p}{R} \partial_\eta T \right) \left(\frac{p}{R} \partial_\eta \mathbf{v} \right)^{-2} \quad (\text{A9})$$

$$R_{iv} = \left(\frac{T_v}{c_p} - \frac{p}{R} \partial_\eta T_v \right) \left(\frac{p}{R} \partial_\eta \mathbf{v} \right)^{-2} \quad (\text{A10})$$

$$\tilde{R}_i = \begin{cases} R_{iv} & \text{for } R_{iv} < R_i \\ R_i & \text{else.} \end{cases} \quad (\text{A11})$$

In Equation A10, the buoyancy frequency is computed using the virtual temperature, $T_v = T (1 + 0.61q)$ (q is the water vapor mass mixing ratio) to take the effect of water vapor content on static stability in the troposphere into account (e.g., Holtslag & Boville, 1993). The dependencies of the diffusion coefficients on the Richardson number are introduced so that the diffusion scheme responds to dynamic instability of the resolved flow. This is particularly relevant for the horizontal diffusion scheme in GW-resolving simulations (Becker, 2009). The offset Richardson number in Equations A6 and A7, R_{i0} , is 0 in the lower troposphere but is increased to a value of 0.25 from the upper troposphere to the midstratosphere (between $\eta = 0.2$ and 0.02). This measure follows the linear dynamic instability criterion by Miles (1961) and Howard (1961) for GWs. The horizontal mixing length, l_h , is prescribed as a function of η . The vertical mixing length includes Prandtl's formula (Holtslag & Boville, 1993):

$$l_z = \left(\frac{1}{k_a (z - z_s)} + \frac{1}{l_{asym}(\eta)} \right)^{-1}. \quad (\text{A12})$$

Here, $k_a = 0.4$ is the van Kàrman constant, z_s is the geometric height of the orography above sea level, and l_{asym} is an asymptotic mixing length that is prescribed as a function of η . The mixing length profiles are included in Figures A1 and A2.

As noted by Gassmann (2018), the assumption of a fixed Prandtl number for turbulent vertical diffusion may violate the second law when the stratification is stable but the temperature decreases with altitude. In the HIAMCM we apply a new constraint for P_r in Equation 2 that is based on the quasi-stationary SGS turbulent kinetic energy equation (e.g., Becker, 2004, his Equation (5)):

$$0 = K_h |\mathbf{S}_h|^2 + K_f (\mathbf{S}_{hf} \nabla) \cdot \mathbf{v} + K_z (\partial_z \mathbf{v})^2 - \frac{K_z}{P_r} N^2 - \epsilon_{mech}. \quad (\text{A13})$$

Here, $\epsilon_{mech} \geq 0$ is the mechanical dissipation due to molecular viscosity. A conservative constraint for P_r is obtained from Equation A13 when neglecting the shear production rates from horizontal momentum diffusion, yielding $K_z (\partial_z \mathbf{v})^2 - K_z N^2 / P_r \geq 0$. This inequality is satisfied by

$$P_r = \max(1, R_i). \quad (\text{A14})$$

This constraint ensures that the entropy production from mechanical and thermal dissipation due to turbulent vertical momentum and heat diffusion is positive definite:

$$K_z (\partial_z \mathbf{v})^2 + c_p K_z (T P_r)^{-1} (g/c_p + \partial_z T) \partial_z T \geq 0. \quad (\text{A15})$$

An approximate formula for the dynamic molecular viscosity based on the original paper of Dalgarno and Smith (1962) was given by (Vadas, 2007, her Equation 2). According to a more recent study by Vadas and Crowley (2017), this formula does not agree with some data for vertical diffusion above altitudes of $z \sim 220$ km. They argued that this was likely because the free path of atomic oxygen (the major constituent at these altitudes) becomes comparable to the density scale height there and therefore that vertical shears of horizontal velocities are not easily smoothed out by molecular viscosity. Vadas and Crowley (2017) suggested replacing the original formula for molecular dynamic viscosity by an effective dynamic viscosity that decreases rapidly with altitude for $z \geq 220$ km. Vadas et al. (2019) noted that this effect is observed in GW measurements since λ_z stops increasing exponentially with altitude at about $z \geq 220$ km (e.g., Nicolls et al., 2014). Here we account for this new result by writing the kinematic viscosity (dynamic viscosity divided by density) used in the vertical diffusion scheme as

$$\nu_z = (\nu_1^{-1} + \nu_{z,max}^{-1})^{-1}, \quad \nu_{z,max} = 3.5 \times 10^6 \text{ m}^2 \text{ s}^{-1} \quad (\text{A16})$$

$$\nu_1 = 3.34 \times 10^{-7} \left(\frac{T}{K} \right)^{1.71} \left(\frac{R}{J K^{-1} \text{kg}^{-1}} \right) \left(\frac{p}{Pa} \right)^{-1} \text{ m}^2 \text{ s}^{-1}, \quad (\text{A17})$$

where ν_1 corresponds to Equation 2 in Vadas (2007) and $\nu_{z,max}$ is an upper limit for ν_z . Since the density does not change notably over horizontal distances comparable with the free path of atomic oxygen in the upper thermosphere, thereby implying that molecular viscosity can still effectively smooth out horizontal shears, we apply almost the full molecular viscosity in the horizontal diffusion scheme:

$$\nu_h = (\nu_1^{-1} + \nu_{h,max}^{-1})^{-1}, \quad \nu_{h,max} = 3 \times 10^8 \text{ m}^2 \text{ s}^{-1}. \quad (\text{A18})$$

The surface drag coefficients, C_D for momentum and sensible heat and C_{Dq} for specific humidity, are defined according to the local boundary layer scheme described by Holtslag and Boville (1993):

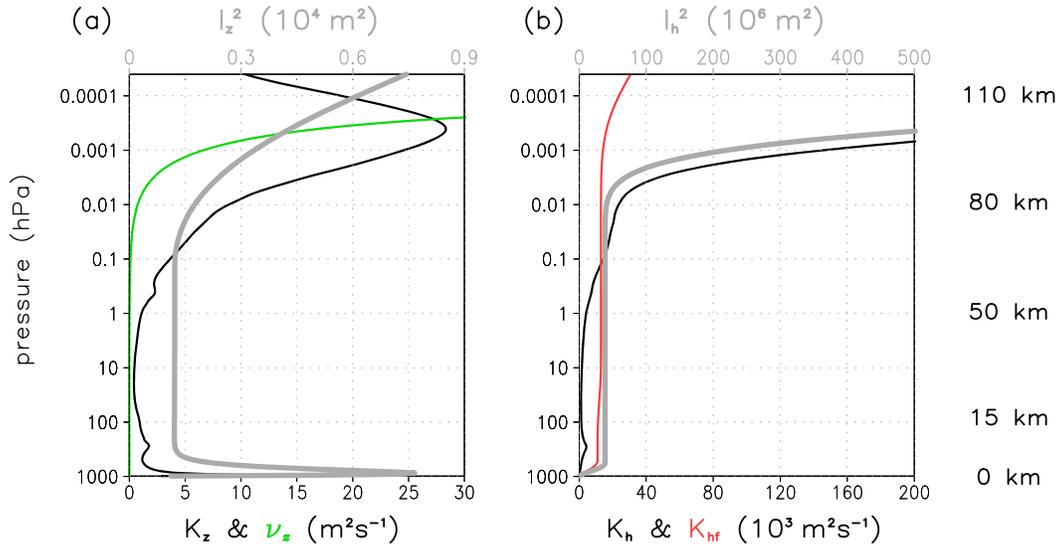


Figure A1. (a) Squared vertical mixing length (gray curve, upper axis) and globally averaged turbulent and molecular vertical diffusion coefficients (black and green curves, respectively, lower axis) in the lower and middle atmosphere. (b) Squared horizontal mixing length (gray curve, upper axis), globally averaged turbulent horizontal diffusion coefficient (black curve, lower axis), and prescribed horizontal diffusion coefficient used for hyperdiffusion (red curve, lower axis). The vertical coordinate is the model's hybrid vertical coordinate times 1,013 hPa. Approximate geometric altitudes are given on the right-hand side of panel (b).

$$C_D = C_N |\mathbf{v}_L| F_L(R_{iL}), \quad c_N = \left(\frac{k_a}{\ln((z_L + z_r)/z_r)} \right)^2 \quad (\text{A19})$$

$$F_L(R_{iL}) = \begin{cases} 1 - 9 R_{iL} \left(1 + 75 c_N \sqrt{|R_{iL}|(z_L + z_r)/z_r} \right)^{-1} & \text{for } R_{iL} < 0 \\ (1 + 9 R_{iL})^{-1} & \text{for } R_{iL} \geq 0. \end{cases} \quad (\text{A20})$$

$$R_{iL} = \frac{g}{T_L(1 + 0.61 q_p)} \left(\frac{g}{c_p} + \frac{(T_L - T_s)(1 + 0.61 q_L) + 0.61 T_L (q_L - q_s)}{z_L} \right) \times (\mathbf{v}_L^2 / z_L^2)^{-1} \quad (\text{A21})$$

$$C_{Dq} = 1.25 \times C_D. \quad (\text{A22})$$

Here, the surface drag coefficient for specific humidity is scaled by a factor of 1.25 relative to momentum and sensible heat to tune the intensity of the tropospheric moisture cycle. Furthermore, \mathbf{v}_L , T_L , and q_L are the horizontal wind, temperature, and specific humidity at the lowest full model layer (which is assumed as the top of the Prandtl layer), and T_s and q_s denote the surface temperature and surface specific humidity (see also Equations 9–11 in section 2). The height of the lowest full model layer above the orography is denoted by z_L , and z_r is the roughness length. We use $z_r = 0.005$ m over ocean surfaces and $z_r = 0.05$ m over land and ice surfaces.

Figures A1 and A2 show the mixing lengths, the global mean turbulent and molecular viscosities, and the prescribed hyperdiffusion coefficient. Figure A3 shows the global-mean mechanical dissipation rates in Equation 2. All profiles refer to the control simulation for January. Several features are worth mentioning: (1) Both turbulent viscosities become large in the MLT, which is triggered by instability of resolved GWs causing small Richardson numbers (Becker, 2009; Lund & Fritts, 2012). (2) The turbulent vertical diffusion coefficient (black curves in Figures A1a and A2a) becomes irrelevant compared to the molecular vertical viscosity (green curves in Figures A1a and A2b) above about ~ 0.0003 hPa ($z \sim 100$ km). The turbulent horizontal diffusion coefficient is exceeded by the horizontal molecular viscosity above $\sim 5 \times 10^{-7}$ hPa ($z \sim 230$

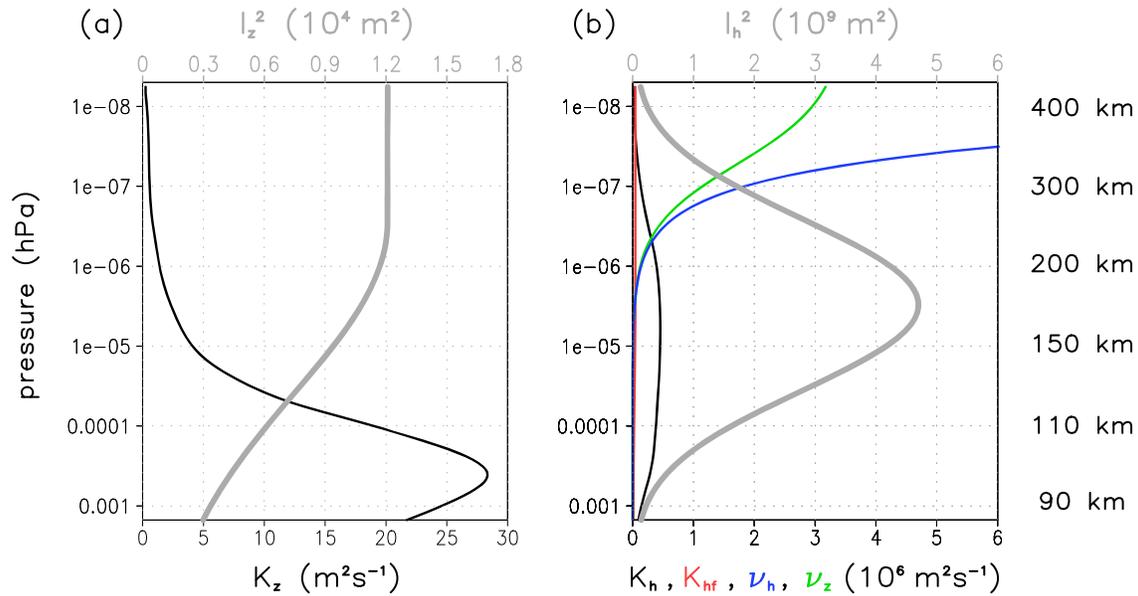


Figure A2. Same as Figure A1 but for the thermosphere. Panel (b) includes the globally averaged molecular viscosities used for vertical and horizontal diffusion (green and blue curves, respectively, lower axis).

km, black and blue curves in Figure A2b). (3) The dissipation rate from hyperdiffusion (red curves in Figure A3) is negligible compare to the dissipation rates from turbulent and molecular viscosities at all altitudes (black and green solid curves in Figure A3). (4) The dissipation from macroturbulent horizontal and vertical momentum diffusion is mainly due to small scales (horizontal wavelengths shorter than 2,000 km, cf. the solid to the dashed curves in Figure A3a). (5) Horizontal momentum diffusion from molecular viscosity mainly dissipates the small scales, whereas molecular vertical momentum diffusion mainly dissipates large horizontal scales (horizontal wavelengths larger than 2,000 km, cf. the solid to the dashed curves in Figure A3b). (6) Ion drag is the predominant dissipation mechanism above ~ 200 km (blue curve in Figure A3b). The corresponding dissipation is mainly due to the large scales (not shown).

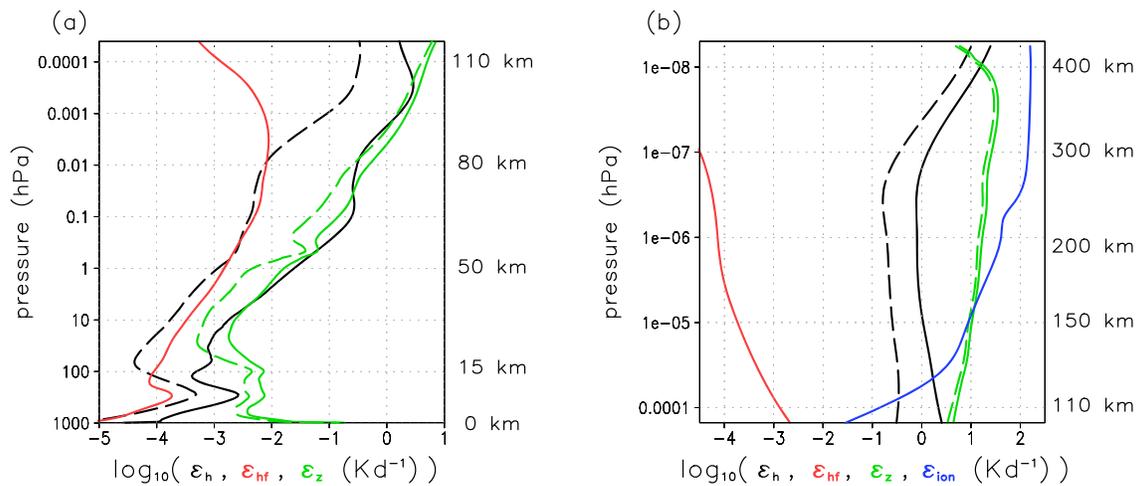


Figure A3. Global mean mechanical dissipation rates during January from different processes (see Equations 1 and 2). Black curves: turbulent plus molecular horizontal momentum diffusion; red curves: hyperdiffusion of momentum; green curves: turbulent plus molecular vertical momentum diffusion; blue curve: ion drag. Solid curves show the contributions from horizontal wavenumbers ($n \leq 240$), while dashed curves show the corresponding contributions from the large scales only ($n \leq 20$). The vertical coordinate is the model's hybrid vertical coordinate times 1,013 hPa. Approximate geometric altitudes are given on the right-hand side of each panel.

Appendix B: Nonhydrostatic Correction

In the following we derive a nonhydrostatic correction of the geopotential, Φ^{nh} , which is added to the horizontal momentum equation of the HIAMCM (see Equation 1). Our method follows Klingbeil and Burchard (2013) who extended a hydrostatic estuarine circulation model by including nonhydrostatic effects. Similar methods were also developed for atmospheric circulation models (e.g., Janjic et al., 2001).

Neglecting viscosity in the vertical momentum equation with z as vertical coordinate, the material vertical acceleration, \dot{w} , is due to the nonhydrostatic pressure force (the gravity acceleration is canceled by the hydrostatic pressure force). In the vertical coordinate system of the HIAMCM, the nonhydrostatic vertical pressure force can be expressed in terms of the nonhydrostatic geopotential as $-\partial_\eta \Phi^{nh} / \partial_\eta z$, hence,

$$\dot{w} = -\partial_\eta \Phi^{nh} / \partial_\eta z. \quad (\text{B1})$$

We assume that nonhydrostatic effects are not relevant in the troposphere. Therefore, Φ^{nh} is computed only for altitudes where $b(\eta) = 0$, which is equivalent to $p = a(\eta)$ (see section 2). We use $b = 0$ for $\eta < \eta_1 = 0.09$, corresponding to an altitude of $z \sim 15$ km above which nonhydrostatic effects are included. We furthermore use (see also Equation 8 in section 2)

$$w \approx -\frac{RT}{pg} \dot{\eta} \partial_\eta p = -\frac{RT}{pg} \dot{p} = \frac{RT}{pg} \int_0^\eta \vartheta \partial_{\bar{\eta}} p \, d\bar{\eta} \text{ for } \eta < \eta_1, \quad (\text{B2})$$

where $\vartheta = \nabla \cdot \mathbf{v}$. Combining Equations B2 and B1 leads to

$$\Phi^{nh} = -\int_\eta^{\eta_1} \frac{RT}{pg} \dot{w} \partial_{\bar{\eta}} p \, d\bar{\eta} \approx \int_\eta^{\eta_1} \left(\frac{RT}{pg}\right)^2 \dot{p} \partial_{\bar{\eta}} p \, d\bar{\eta} \text{ or} \quad (\text{B3})$$

$$\Phi^{nh} = -\int_\eta^{\eta_1} \left(\frac{RT}{pg}\right)^2 \left(\dot{p} \vartheta + \int_0^{\bar{\eta}} \partial_t \vartheta \partial_{\bar{\eta}} p \, d\bar{\eta} + \mathbf{v} \cdot \int_0^{\bar{\eta}} \nabla \vartheta \partial_{\bar{\eta}} p \, d\bar{\eta} \right) \partial_{\bar{\eta}} p \, d\bar{\eta} \quad (\text{B4})$$

for $\eta < \eta_1$, which is used in the HIAMCM. Here, the time derivative (tendency) of the horizontal divergence, $\partial_t \vartheta$, is computed from the previous time step.

Appendix C: Moist Convection and Large-Scale Condensational Heating

The moisture cycle in the HIAMCM employs a simplified convection scheme that relaxes the temperature and specific humidity to reference values that correspond to piecewise moist-adiabatic profiles (depending on the actual temperature and water vapor content in the column). To define these profiles, we recall that moist convection and nonconvective condensation take place in only limited areas compared to a grid box of a GCM. Hence, condensation or convection will take place before the saturation or moist instability level is met. Furthermore, convective updrafts are surrounded by compensating subsidence of water vapor, as well as by convective downdrafts with reevaporation of cloud water. To roughly represent these effects, we specify the reference specific humidity as

$$q_r = \begin{cases} q_{sat}(p, T) r(T) & \text{if } q_{sat}(p, T) r(T) < q \\ q & \text{else,} \end{cases} \quad (\text{C1})$$

where $r(T)$ is defined as

$$r(T) = \begin{cases} 0.83 & \text{if } T > T_2 = 285 \text{ K} \\ 0.71 & \text{if } T < T_1 = 220 \text{ K} \\ 0.83 - 0.12 \cos^2\left(\frac{\pi}{2} \frac{T - T_1}{T_2 - T_1}\right) & \text{else.} \end{cases} \quad (\text{C2})$$

The saturation water vapor mass mixing ratio is given by the integrated Clausius-Clapeyron equation:

$$q_{sat}(T, p) = \frac{R}{R_v} \frac{p_{ref}}{p} \exp\left(\frac{\ell}{R_v} \left(\frac{1}{T_{ref}} - \frac{1}{T}\right)\right). \quad (C3)$$

Here, $R = 287 \text{ m}^2 \text{ s}^{-2} \text{ K}^{-1}$ and $R_v = 461 \text{ m}^2 \text{ s}^{-2} \text{ K}^{-1}$ denote the gas constants of dry air (in the lower atmosphere) and water vapor, respectively; $p_{ref} = 6.11 \text{ hPa}$ and $T_{ref} = 273 \text{ K}$ are reference values, and $\ell = 2.5 \times 10^6 \text{ m}^2 \text{ s}^{-2}$ is the latent heat of water vapor per unit mass. The surface humidity, q_s , is defined as

$$q_s = f_s q_{sat}(T_s, p_s), \quad (C4)$$

where $f_s = f_s(\lambda, \phi)$ is a prescribed relative surface humidity. The reference temperature, T_r , used in the convection scheme is computed piecewise using

$$T_r = T_1 - \frac{\Phi - \Phi_1}{c_p} - \frac{\ell (q_r - q_{r1})}{c_p} \text{ for } T_r \geq T. \quad (C5)$$

We start with $T_1 = T_L$, where T_L is the temperature at the lowest full model layer, then compute T_r at greater heights from Equation C5 until the condition $T_r < T$ is met; the temperature T at this level is then defined as the new T_1 from which the computation of T_r according to Equation C5 is continued, and so forth; the computation of T_r is terminated at $\eta = 0.09$ (corresponding to $\sim 90 \text{ hPa}$), which is the uppermost level at which moist convection and large-scale condensation are considered.

The relaxation rates used in the moist convection scheme follow the mass flux method of Kuo (1965): The moisture subject to convective condensation corresponds to that transported into the air column by advection and diffusion. Moist convection is assumed to reach up to the highest level where the actual temperature falls below the reference temperature. Denoting this level by the symbol η_c , the moisture supply into the convective column is

$$\dot{M} = \int_{\eta_c}^1 \left(-\mathbf{v} \cdot \nabla q - \eta \partial_\eta q + \frac{1}{\partial_\eta p} \nabla \cdot (K_h \partial_\eta p \nabla q) + \frac{g}{\partial_\eta p} \partial_\eta \left(\frac{\rho^2 g}{\partial_\eta p} K_z \partial_\eta q \right) \right) \frac{\partial_\eta p}{g} d\eta. \quad (C6)$$

The corresponding convective condensation and heating rates in Equations 3 and 2 are formulated as relaxation:

$$Q_{conv} = c_p \alpha_T (T_r - T) \quad (C7)$$

$$C_{conv} = \alpha_q (q_r - q), \quad (C8)$$

where the relaxation rates are determined from mass and energy conservation:

$$\alpha_T = \frac{\ell}{c_p} \dot{M} \left(\int_{\eta_c}^1 (T_r - T) \frac{\partial_\eta p}{g} d\eta \right)^{-1} \quad (C9)$$

$$\alpha_q = \dot{M} \left(\int_{\eta_c}^1 (q_r - q) \frac{\partial_\eta p}{g} d\eta \right)^{-1}. \quad (C10)$$

Between η_c and $\eta_0 = 0.09$ (corresponding to $\sim 90 \text{ hPa}$) large-scale condensation in terms of relaxation toward the reference specific humidity is applied. This yields the following condensation and latent heating rates used in Equations 2 and 3 in section 2:

$$C_{large} = \begin{cases} -\alpha_{large} (q_r - q) & \text{if } \eta_c < \eta < \eta_0 \text{ and } q > q_r \\ 0 & \text{else} \end{cases} \quad (C11)$$

$$Q_{large} = \ell C_{large}. \quad (C12)$$

Here, the relaxation rate is $\alpha_{large} = (30 \text{ min})^{-1}$. In addition, the water vapor mixing is relaxed to 0 with a time constant of 365 days everywhere, ensuring reasonable water vapor mixing ratios above the tropopause.

Appendix D: Thermodynamic Relations for Variable Gas Constant

The assumption of a fixed gas constant for the dry phase of air is appropriate in the lower and middle atmosphere, where N_2 and O_2 are dominant and well mixed. In the lower thermosphere, however, N_2 , O_2 , and O are approximately equally dominant and are demixed (have different density scale heights). At about 200 km $\leq z \leq$ 400 km, O is dominant. At higher altitudes, He and then H become the dominant constituents. Thus, atomic oxygen is the major constituent in the F region. The overall result is a significant increase of the resultant gas constant, R , and of the specific heat per unit mass at constant density, c_v , with increasing height (e.g., Vadas, 2007). In order to precisely simulate the full dynamics of the variable mixture of the neutral gas components in the thermosphere, one has to apply multiphase hydrodynamics (e.g., Del Genio et al., 1979). On the other hand, a dynamical core of a GCM usually solves fluid dynamical equations for a single phase. Hence, it is an issue to extend the dynamical core of a GCM into the thermosphere. Different approaches are applied in different thermosphere-ionosphere models (e.g., H.-L. Liu et al., 2018). The HIAMCM uses sensible heat as prognostic thermodynamic variable, as is usual in many GCMs. In the following we show that in this case the model equations can consistently be extended into the thermosphere by invoking functional dependencies of the form $R = R(p)$ and $c_p = c_p(T)$.

Consider an air parcel in local thermodynamic equilibrium that consists of different components (or phases) denoted by the index n . For each phase, the equation of state is $p_n = R_n \rho_n T$, and the internal energy and sensible heat per unit mass can be written as $e_n = c_{vn} T$ and $h_n = e_n + p_n / \rho_n = c_{pn} T$, respectively. Here c_{vn} and $c_{pn} = c_{vn} + R_n$ are the specific heat capacities per unit mass at constant density and constant pressure for each phase. Defining the mass mixing ratios for each phase as $q_n = \rho_n / \rho$ with $\rho = \sum_n \rho_n$, the equation of state, and the internal energy and sensible heat per unit mass of the air parcel can be written like the corresponding relations for the individual phases:

$$\sum_n p_n = p = \rho R T, \quad R = \sum_n q_n R_n \quad (D1)$$

$$\sum_n q_n e_n = e = c_v T, \quad c_v = \sum_n q_n c_{vn} \quad (D2)$$

$$\sum_n q_n h_n = h = c_p T, \quad c_p = \sum_n q_n c_{pn} = c_v + R. \quad (D3)$$

However, the differentials of the thermodynamic potentials for the air parcel lead to more complicated expressions than for the individual phases. For example, the differential of internal energy reads after a few steps:

$$de = \sum_n (q_n de_n + e_n dq_n) = \frac{p}{\rho^2} dp + T ds + \sum_n g_n dq_n. \quad (D4)$$

Here, $s = \sum_n q_n s_n$ is the entropy per unit mass of the air parcel, s_n is the entropy per unit mass of phase n , and $g_n = h_n - T s_n$ is the Gibbs free energy of phase n . Likewise, the differential of sensible heat of the air parcel becomes

$$dh = \sum_n (q_n dh_n + h_n dq_n) = \frac{dp}{\rho} + T ds + \sum_n g_n dq_n. \quad (D5)$$

Equations D4 and D5 imply that using the internal energy or the sensible heat of the air parcel as a prognostic thermodynamic variable requires the knowledge of $\sum_n g_n dq_n$. It is therefore not straightforward to extend a GCM into the thermosphere and account for the variable gas constant and heat capacities in a thermodynamically consistent way. In the following we propose a simple solution to this problem.

The thermodynamic differential of sensible heat for a single phase,

$$dh = \frac{1}{\rho} dp + T ds = \frac{1}{\rho} dp + T \left(\frac{\partial s}{\partial p} \right)_T dp + T \left(\frac{\partial s}{\partial T} \right)_p dT \quad (D6)$$

yields for the second derivatives

$$\frac{\partial^2 h}{\partial T \partial p} = -\frac{1}{\rho^2} \left(\frac{\partial \rho}{\partial T} \right)_p + \left(\frac{\partial s}{\partial p} \right)_T + T \frac{\partial^2 s}{\partial T \partial p}, \quad \frac{\partial^2 h}{\partial p \partial T} = T \frac{\partial^2 s}{\partial p \partial T}. \quad (D7)$$

These two expressions must be equal; hence,

$$\left(\frac{\partial s}{\partial p} \right)_T = \frac{1}{\rho^2} \left(\frac{\partial \rho}{\partial T} \right)_p. \quad (D8)$$

Using the ideal gas law with $R = R(p)$, this leads to

$$\left(\frac{\partial s}{\partial p} \right)_T = -\frac{R}{p} \text{ for } R = R(p). \quad (D9)$$

The same expression is obtained if R is constant. However, if we allow R to be dependent on both pressure and temperature, the situation would become more complex:

$$\left(\frac{\partial s}{\partial p} \right)_T = -\frac{R}{p} - \frac{T}{p} \left(\frac{\partial R}{\partial T} \right)_p \text{ for } R = R(p, T). \quad (D10)$$

Plugging Equation D9 or D10 back into Equation D6 and substituting $1/\rho$ in the first term on the rhs by the ideal gas law, we get

$$dh = T \left(\frac{\partial s}{\partial T} \right)_p dT \text{ for } R = R(p) \quad (D11)$$

$$dh = T \left(\frac{\partial s}{\partial T} \right)_p dT - \frac{T}{R \rho} \left(\frac{\partial R}{\partial T} \right)_p dp \text{ for } R = R(p, T). \quad (D12)$$

Hence, only for $R = R(p)$ is the differential of the sensible heat proportional to the differential of the temperature:

$$dh = c_p dT \text{ for } R = R(p), \quad (D13)$$

where

$$c_p = T \left(\frac{\partial s}{\partial T} \right)_p \quad (D14)$$

is the heat capacity per unit mass at constant pressure. Because of Equation D13, the sensible heat is a function of the temperature alone and the heat capacity at constant pressure can only be a function of the temperature:

$$c_p = c_p(T) \text{ for } R = R(p). \quad (D15)$$

The implication of this analysis is that by specifying $R = R(p)$ and $c_p = c_p(T)$, a GCM with sensible heat as the prognostic thermodynamic variable can consistently be extended into the thermosphere. Indeed, the equations of motion of the HIAMCM as specified in section 2 are thermodynamically and energetically consistent for the whole model domain for $R = R(p)$ and $c_p = c_p(T)$.

Our specification of $R(p)$ and $c_p(T)$ in the HIAMCM is based on the approximate formulas given in Vadas (2007) for R and $\gamma = c_p/c_v$ based on results from the TIME-GCM. In the following we recapitulate these formulas and show how they are used to construct the functions $R(p)$ and $c_p(T)$. We define a reference temperature, $T_R(p)$, corresponding to active solar conditions as (Profile II in Figure 1 of Vadas, 2007):

$$T_R = \begin{cases} T_1 = 290 \text{ K} & \text{if } p > p_1 = 0.01 \text{ hPa} \\ T_2 = 1000 \text{ K} & \text{if } p < p_2 = 7 \times 10^{-9} \text{ hPa} \\ T_1 + (T_2 - T_1) \sin^2 \left(\frac{\pi}{2} \frac{\ln p_1 - \ln p}{\ln p_1 - \ln p_2} \right) & \text{else.} \end{cases} \quad (D16)$$

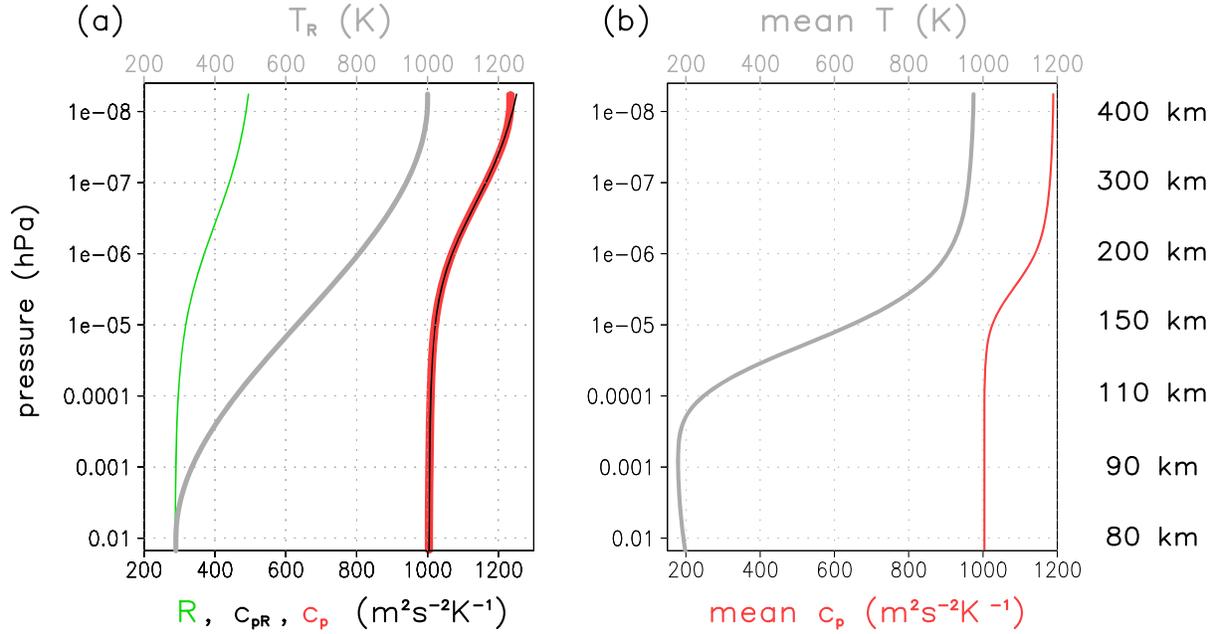


Figure D1. (a) Pressure dependence of the gas constant using Equation D17 (green curve, lower axis) using the reference temperature profile according to Equation D16 (gray curve, upper axis). The black curve (lower axis) shows the reference profile of the heat capacity according to Equation D19, whereas the thick red curve shows the polynomial fit according to Equation D20. (b) Simulated global-mean temperature (upper axis) and heat capacity (lower axis) from Equation D20 during January.

The corresponding profile of $R(p)$ is computed iteratively. We first set $R(p) = R_1 = 286.04 \text{ m}^2 \text{ s}^{-2} \text{ K}^{-1}$ and compute a reference density profile according to $\rho_R(p) = p / (R(p) T_R(p))$. We then use Equation 3 in Vadas (2007) to compute the first iteration of $R(p)$:

$$R(p) = \max(R_1, X_{MW}^{-1} \times 8314.5 \text{ m}^2 \text{ s}^{-2} \text{ K}^{-1})$$

$$X_{MW} = \frac{X_{MW1} - X_{MW2}}{2} \left(1 - \tanh \left(-\frac{\ln(\rho_R(p) / (\text{g m}^{-3})) + 14.9}{4.2} \right) \right) + X_{MW2}. \quad (\text{D17})$$

Here, X_{MW} is the mean molecular or atomic weight of the gas particles, and $X_{MW1} = 28.97$ and $X_{MW2} = 16$ are the tropospheric and thermospheric ($z \sim 400\text{--}450 \text{ km}$) values of X_{MW} , respectively. The result from Equation D17 is then used to update $\rho_R(p)$, which in turn is used to iterate $R(p)$ using Equation D17, and so forth. This procedure converges after only a few iterations. As the next step we use Equation 4 from Vadas (2007) to compute

$$\gamma_R(p) = \frac{1}{2} (\gamma_1 - \gamma_2) \left(1 - \tanh \left(-\frac{\ln(\rho_R(p) / (\text{g m}^{-3})) + 15.1}{4} \right) \right) + \gamma_2 \text{ and} \quad (\text{D18})$$

$$c_{pR}(p) = \max \left(\frac{\gamma_R(p)}{\gamma_R(p) - 1} R(p), c_{p1} \right), \quad (\text{D19})$$

where $\gamma_1 = 1.4004$, $\gamma_2 = 1.67$, and $c_{p1} = 1,004 \text{ m}^2 \text{ s}^{-2} \text{ K}^{-1}$. Since the profile $c_{pR}(p)$ corresponds to the reference temperature profile, $T_R(p)$, these two profiles can be used to construct $c_p(T)$. This is done using a simple fit with Legendre polynomials:

$$c_p(T) = \sum_{n=0}^7 c_n P_n(\zeta) \text{ with } \zeta = \begin{cases} -1 & \text{if } T < T_1 \\ +1 & \text{if } T > T_2 \\ 2(T - T_1) / (T_2 - T_1) - 1 & \text{else.} \end{cases} \quad (\text{D20})$$

For the given parameters of the reference temperature profile, the expansion coefficients in Equation D20 are in units of $\text{m}^2 \text{s}^{-2} \text{K}^{-1}$: $c_0 = 1052.235$, $c_1 = 89.9357$, $c_2 = 62.0863$, $c_3 = 24.8673$, $c_4 = 4.6752$, and $c_5 = 0.1940$. Note that we used a higher reference temperature for the lower and middle atmosphere than in Vadas (2007). This choice ensures that Equation D20 does not yield any notable deviation from c_{p1} below the mesopause. Figure D1 shows $T_R(p)$ (gray curve) and $R(p)$ (green curve), as well as $c_{pR}(p)$ (thin black curve) and $c_p(T)$ (thick red curve) for $T = T_R(p)$. Comparing the two heat capacity profiles confirms that $c_p(T)$ approximates the reference profile very well. The simulated global mean profiles for T and c_p from the control simulation (panel b) confirm that $c_p(T)$ deviates from c_{p1} by up to about 20% at high altitudes. This deviation is even larger in the temperature maxima of the diurnal tide (not shown).

Appendix E: Specification of Ion Drag

In the thermosphere, the gyrofrequency of the ions is much larger than the collision frequency between the ions and the neutrals, ν_{in} . Therefore, the ions can only move easily along the magnetic field lines, and their velocity in that direction is given by the projection of the neutral wind on the direction of the magnetic field lines, defined by the 3-D unit vector \mathbf{b}_3 . The drag of the ions on the neutrals (ion drag) vanishes in the direction of \mathbf{b}_3 but is finite for the wind components perpendicular to \mathbf{b}_3 . Denoting the horizontal components of \mathbf{b}_3 as \mathbf{b} , the resulting ion drag in the horizontal momentum equation of the atmospheric flow can be written as (C. H. Liu & Yeh, 1969)

$$\mathbf{d} = -\nu_{in} \frac{\rho_i}{\rho} (\mathbf{v} - \mathbf{v}_i). \quad (\text{E1})$$

Here, ρ_i is the ion density, and $\mathbf{v}_i = \mathbf{b} (\mathbf{b} \cdot \mathbf{v})$ is the horizontal ion velocity.

In the current version of the HIAMCM we employ the ion-drag parameterization of Hong and Lindzen (1976), which was also used in other GCMs (e.g., Fomichev et al., 2002). This scheme is based on the simplification that the magnetic poles coincide with the geographical poles. In this case, \mathbf{b}_3 has only meridional and vertical components. With I denoting the dip angle of the magnetic field, we can write $\mathbf{b}_3 = \cos I \mathbf{e}_\phi - \sin I \mathbf{e}_z$. Hence, $\mathbf{b} = \cos I \mathbf{e}_\phi$, and the ion drag can be written as

$$\mathbf{d} = -\nu_i (u \mathbf{e}_\lambda + \sin^2 I v \mathbf{e}_\phi), \quad \nu_i = \nu_{in} \frac{\rho_i}{\rho}, \quad (\text{E2})$$

where u and v are the zonal and meridional neutral wind components and ν_i is a damping rate. Within the simplified geometry, the dip angle can be expressed as a function of latitude using $\tan I = 2 \tan \phi$. For their linear tidal wave analysis, Hong and Lindzen (1976) assumed $\nu_i = \nu_i(z)$ based on averaging over typical dayside and nightside ion densities, and they provided the following parametric form

$$\nu_i(z) = 5 \times 10^{-10} \text{ s}^{-1} \sum_{i=1}^3 A_i \exp(a_i (1 - r_i(z) - \exp(-r_i(z)))). \quad (\text{E3})$$

Here, the coefficients a_i and A_i and the function $r_i(z)$ are for solar minimum conditions:

$$\begin{aligned} A_1 &= 6.6 \times 10^4, \quad a_1 = 1.4, \quad r_1 = \frac{z - 150 \text{ km}}{0.2 z + 1 \text{ km}} \\ A_2 &= 1.56 \times 10^5, \quad a_2 = 1.0, \quad r_2 = \frac{z - 225 \text{ km}}{42 \text{ km}} \\ A_3 &= 3.0 \times 10^5, \quad a_3 = 0.35, \quad r_3 = \frac{z - 275 \text{ km}}{0.1 z + 1 \text{ km}} \end{aligned} \quad (\text{E4})$$

and for solar maximum conditions

$$\begin{aligned} A_1 &= 1.15 \times 10^5, \quad a_1 = 1.4, \quad r_1 = \frac{z - 150 \text{ km}}{0.2 z + 1 \text{ km}} \\ A_2 &= 2.75 \times 10^5, \quad a_2 = 1.0, \quad r_2 = \frac{z - 240 \text{ km}}{52 \text{ km}} \\ A_3 &= 1.05 \times 10^6, \quad a_3 = 0.2, \quad r_3 = \frac{z - 300 \text{ km}}{0.1 z + 1 \text{ km}} \end{aligned} \quad (\text{E5})$$

For the simulations in this study, we applied this ion drag scheme for solar maximum conditions in the following way. At each time step we compute the geometric height of the global-mean temperature, T , according to

$$z = \tilde{z} \frac{\tilde{z} + a_e}{a_e}, \quad \tilde{z} = g^{-1} \int_p^{p_{00}} \frac{R(\tilde{p}) T(\tilde{p})}{\tilde{p}} d\tilde{p}, \quad (\text{E6})$$

where a_e is the Earth radius and $p_{00} = 1,013$ hPa. We then compute the damping rate as a function of model layer (pressure in the thermosphere) from Equations E3 and E5. A dependence of the ion density on the daily cycle is introduced using

$$\nu_i \rightarrow \nu_i \times (0.55 + 1.24 \cos\Theta_o) \quad (\text{E7})$$

where Θ_o is the zenith angle of the solar insolation.

Appendix F: Radiation Scheme

In the following we describe idealized computation of radiative energy flux densities and associated heating rates as implemented in the HIAMCM. These methods represent a further development of the previous radiation scheme of the KMCM as described in Knöpfel and Becker (2011) and Becker et al. (2015).

F1. Shortwave Radiation

With S_o denoting the solar constant, we divide the solar insolation energetically into five bands according to

$$S_o = \sum_{k=1}^5 S_o^k \quad \text{with} \quad S_o^k = \beta_k S_o. \quad (\text{F1})$$

Here, the β_k describe the fractions of the solar insolation that are subject to absorption by ozone ($\beta_1 = 0.0077$ for UV-C, $\beta_2 = 0.225$ for UV-A and UV-B) and by water vapor in the visible regime ($\beta_3 = 0.185$). Furthermore, the fraction $\beta_4 = 7.6 \times 10^{-6}$ is absorbed in the thermosphere, and $\beta_5 = 1 - \sum_{k=1}^4 \beta_k$ is subject to absorption by tropospheric clouds. The eccentricity of the Earth's orbit around the Sun is accounted for by specifying the solar constant as

$$S_o = 1,362 \text{ W m}^{-2} + \cos(2\pi\tau) \times 55 \text{ W m}^{-2}, \quad (\text{F2})$$

where τ denotes time in units of years. Using pressure p as vertical coordinate, the fluxes in the individual bands are computed from Beer-Bouger-Lambert's law:

$$\cos \Theta_o \frac{\partial S^k}{\partial p} = -g^{-1} (\zeta_k \Lambda_k + \zeta_5 r) S_k \quad \text{with} \quad S^k(p=0) = \cos \Theta_o S_o^k. \quad (\text{F3})$$

Here, $\Theta_o(\lambda, \phi, t)$ is the zenith angle of solar insolation, where λ is longitude, ϕ is latitude, and t is universal time. The mass mixing ratios of the absorber gases are $\zeta_1 = \zeta_2 = q_{O3}$ for ozone, $\zeta_3 = q$ for water vapor, $\zeta_4 = (R_2 - 0.7 R(p) - 0.3 R_1) / (R_2 - R_1)$ (with $R_2 = 570 \text{ m}^2 \text{ s}^{-2} \text{ K}^{-1}$ and $R_1 = 286 \text{ m}^2 \text{ s}^{-2} \text{ K}^{-1}$) as an empirical representation of the mixture of air molecules and atomic oxygen in the thermosphere that give rise to heating due to absorption of EUV radiation. Furthermore, ζ_5 is a proxy for tropospheric clouds. We use the relative humidity for this purpose but multiplied with a function f_c that is unity between 800 and 300 hPa and approaches 0 from 300 to 100 hPa and a value of 0.3 below 800 hPa; hence $\zeta_5 = f_c q / q_{sat}$. The reflection coefficient due to clouds is denoted by the symbol r in Equation F3. The shortwave heating rates are computed according to

$$Q_{SW}^k = -g \frac{\zeta_k \Lambda_k}{\zeta_k \Lambda_k + \zeta_5 r} \partial_p S_k, \quad (\text{F4})$$

and the reflected upward solar energy flux densities fulfill

$$\frac{\partial S_{or}^k}{\partial p} = -g^{-1} \frac{\zeta_5 r}{\zeta_k \Lambda_k} Q_{SW}^k \quad \text{with} \quad S_{or}^k(p = p_s) = \alpha S_o^k(p = p_s). \quad (\text{F5})$$

Here, α is a prescribed surface albedo having values of 0.04, 0.26, and 0.56 for ocean, land, and ice surfaces, respectively. The band strengths, Λ_k , and the reflections coefficient are tunable parameters. In the present model version we use the following values in units of $\text{m}^2 \text{kg}^{-1}$: $\Lambda_1 = 3, 820, \Lambda_2 = 14.6, \Lambda_3 = 0.125, \Lambda_4 = 1.1 \times 10^4, \Lambda_5 = 0.003$, and $r = 8.85 \times 10^{-4}$. Global-mean profiles of constituents and heating rates are presented in section F3.

F2. Longwave Radiation

Computation of the thermal radiation is based on the radiative transfer equation (e.g., Thomas & Stamnes, 2002):

$$\frac{dI_\nu}{ds} = \rho \kappa_\nu \left(-I_\nu + \varpi_\nu \int I_\nu d\Omega (4\pi)^{-1} + (1 - \varpi_\nu) B_\nu \right). \quad (\text{F6})$$

Here, $d\Omega$ is the solid angle element, I_ν is the intensity, κ_ν is the mass extinction coefficient, and B_ν is the Planck function:

$$B_\nu = \frac{2 h \nu^3 / c^2}{e^{h \nu / k_B T} - 1}, \quad (\text{F7})$$

where h is the Planck constant, k_B the Boltzman constant, and ν is the frequency in s^{-1} . The single-scattering albedo, $0 \leq \varpi_\nu \leq 1$, represents the effect of non-LTE. Also note that

$$\pi \int_{\nu=0}^{\infty} B_\nu d\nu = \sigma T^4, \quad (\text{F8})$$

where σ is the Stefan-Boltzmann constant.

We assume a plane-parallel and horizontally isotropic atmosphere, and we apply the Eddington approximation to parameterize the dependence of the intensity on the zenith angle, $\tilde{\vartheta}$. More specifically, I_ν is assumed to be independent of the horizontal direction, and its dependence on $\tilde{\vartheta}$ is parameterized using

$$I_\nu = \frac{U_\nu + D_\nu}{2\pi} + \frac{3(U_\nu - D_\nu)}{4\pi} \cos \tilde{\vartheta}. \quad (\text{F9})$$

Here, U_ν and D_ν are the net upward and downward spectral energy flux densities; that is, U_ν ($-D_\nu$) is the solid-angle integral of $I_\nu \cos \tilde{\vartheta}$ over the upper (lower) half sphere. Multiplying Equation F6 with $\cos \tilde{\vartheta}$, using $dz = ds \cos \tilde{\vartheta}$, integrating over the upper and lower half sphere, and using pressure as vertical coordinate by means of the hydrostatic formula yields the following transfer equations for the upward and downward spectral energy flux densities (Knöpfel & Becker, 2011):

$$\frac{\partial U_\nu}{\partial p} = \frac{\kappa_\nu}{g} \left(\left(\frac{7}{4} - \varpi_\nu \right) U_\nu + \left(\frac{1}{4} - \varpi_\nu \right) D_\nu - (2 - 2\varpi_\nu) \pi B_\nu \right) \quad (\text{F10})$$

$$\frac{\partial D_\nu}{\partial p} = -\frac{\kappa_\nu}{g} \left(\left(\frac{7}{4} - \varpi_\nu \right) D_\nu + \left(\frac{1}{4} - \varpi_\nu \right) U_\nu - (2 - 2\varpi_\nu) \pi B_\nu \right). \quad (\text{F11})$$

We now introduce six frequency bands: the 9.6 μm band of ozone (denoted by the index $k = 1$), the 6.3 μm water vapor band ($k = 2$), the water vapor continuum ($k = 3$), and three compartments of the 15 μm band of carbon dioxide ($k = 4..6$). An additional longwave extinction coefficient corresponds to an absorber in the gray limit (index c). We associate this additional gray limit absorber with the longwave radiative effects of tropospheric clouds and of NO in the mesopause region and lower thermosphere.

We use the decompositions

$$\kappa_\nu = \kappa^k + \kappa'_\nu, \quad \kappa^k = \frac{1}{\nu_{2k} - \nu_{1k}} \int_{\nu_{1k}}^{\nu_{2k}} \kappa_\nu d\nu \quad (\text{F12})$$

$$(\nu_{2k} - \nu_{1k}) X_\nu = X^k + X'_\nu, \quad X^k = \int_{\nu_{1k}}^{\nu_{2k}} X_\nu d\nu, \quad (\text{F13})$$

which are valid for each frequency band (i.e., for $\nu_{1k} \leq \nu \leq \nu_{2k}$) and where X represents either U , D , or B . Furthermore, ϖ_ν is assumed to be independent of frequency for each band; hence, $\varpi_\nu = \varpi^k$ (defined at the end of this section). With these definitions we integrate Equations F10 and F11 over frequency. Due to the strong variations of κ_ν , U_ν , and D_ν with frequency, the frequency-integrated radiative transfer equations include covariance terms of the form $\overline{\kappa'_\nu D'_\nu}$ and $\overline{\kappa'_\nu U'_\nu}$. These terms become important when the frequency band consists of sharp instead of broad lines. In particular, the saturation in the CO₂ 15 μm band is strongly dependent on the reduction of line broadening in the upper troposphere (Pierrehumbert, 2011). Covariance terms of the form $\overline{\kappa'_\nu B'_\nu}$ are negligible because the Planck function varies only slowly with frequency.

Since water vapor is mainly relevant in the troposphere, while the longwave radiative transfer of ozone is important only in the stratosphere and lower mesosphere, and since saturation effects for these absorbers are not important, we neglect the covariance terms for these bands and solve the following transfer equations for $k = 1-3$:

$$\frac{\partial U^k}{\partial p} = \frac{7}{4} \frac{\omega^k}{g} (\kappa^k + \kappa^c) U^k + \frac{1}{4} \frac{\omega^k}{g} (\kappa^k + \kappa^c) D^k - \frac{2 - 2\omega^k}{g} (\kappa^k + \kappa^c) \pi B^k \quad (\text{F14})$$

$$\frac{\partial D^k}{\partial p} = -\frac{7}{4} \frac{\omega^k}{g} (\kappa^k + \kappa^c) D^k - \frac{1}{4} \frac{\omega^k}{g} (\kappa^k + \kappa^c) U^k + \frac{2 - 2\omega^k}{g} (\kappa^k + \kappa^c) \pi B^k. \quad (\text{F15})$$

Here, $\kappa^k = S^k \zeta_k$, where the band strengths, S^k , are tunable parameters and ζ_k represents the mass mixing ratio of ozone ($k = 1$) and water vapor ($k = 2$ and $k = 3$): $\zeta_1 = q_{\text{O}_3}$ and $\zeta_2 = \zeta_3 = q$. The band strengths and frequency ranges applied in the HIAMCM are given in the first three rows of Table 1.

The longwave effects of tropospheric clouds and NO are represented in a simplistic fashion using

$$\kappa^c = S_c f_c q / q_{\text{sat}} + S_{\text{NO}} q_{\text{NO}}. \quad (\text{F16})$$

Here, $f_c q / q_{\text{sat}}$ is the cloud proxy defined in Appendix F1, $S_c = 0.0092$, $S_{\text{NO}} = 0.2$, and the prescribed vertical profile of the assumed mass mixing ratio q_{NO} is given in Figure F1a (see Appendix F3).

The CO₂ 15 μm band is split into three wings (frequency compartments, $k = 4..6$). These wings represent the center of the band that has the strongest lines ($k = 5$), as well the two wings next to the center at lower frequencies ($k = 4$) and at higher frequencies ($k = 6$). As discussed by Pierrehumbert (2011, see also his Figure 2), the extinction coefficient varies strongly with frequency within each wing. The CO₂ band includes further wings, which are orders of magnitude weaker and therefore not very important, even though they are relevant for the climate sensitivity to multiple CO₂ doubling (Pierrehumbert, 2011). Since the infrared radiative transfer calculations in the HIAMCM extend continuously from the surface to the model top, and since CO₂ is relevant at all altitudes up to the lower thermosphere, we must include the frequency variations of the extinction coefficient within each wing $k = 4..6$ to compute the frequency-averaged upward and downward spectral energy flux densities with sufficient accuracy. We found that the method proposed in Knöpfel and Becker (2011) was not sufficient in this respect. In the following we propose a pseudo line-by-line integration method.

A close inspection of the individual line strengths of the CO₂ 15 μm band based on the HITRAN data base (Rothman et al., 2013) showed that each wing $k = 4..6$ consists of two categories of lines, namely, very weak lines and strong lines. The latter mainly account for the frequency-averaged absorption coefficient in each

wing. For these strong lines we use the HITRAN data base to compute (1) the extinction coefficient as a function of frequency and (2) the parameters that determine the line widths as functions of pressure and temperature. We then compute averaged values for these parameters. More specifically, each line is parameterized with the Voigt profile, which includes Doppler and pressure broadening. The half-line width for a Voigt profile is given by (e.g., Thomas & Stamnes, 2002)

$$\gamma_V = 0.5346 \gamma_L + \sqrt{0.2166 \gamma_L^2 + \gamma_G^2} \quad (\text{F17})$$

$$\gamma_L = \left(\frac{T_0}{T}\right)^{n_k} (\gamma_{air} p_{00}^{-1} (p - p_{CO_2}) + \gamma_{CO_2} p_{00}^{-1} p_{CO_2}) \quad (\text{F18})$$

$$\gamma_G = \nu_0 \left(\frac{2 (\ln 2) k_B T}{c^2 m_{CO_2}}\right)^{1/2} \quad (\text{F19})$$

where γ_L and γ_G are the line widths due to pressure broadening (with respect to a Lorentz line profile) and Doppler broadening (with respect to a Gaussian line profile), respectively. The unit in these equations is cm^{-1} , as is convenient in radiative transfer computations. Furthermore, $T_0 = 296 \text{ K}$, $p_{00} = 1,013 \text{ hPa}$, p_{CO_2} is the partial pressure of CO_2 , m_{CO_2} is the mass of the CO_2 molecule in kg, and c is the speed of light. The Lorentz line width, γ_L , comprises the pressure broadening from collisions with both air molecules and CO_2 molecules. Other parameters used in Equations F17–F19 are given in rows 4–6 of Table 1. Using these numbers, Equations F17–F19 yield the mean half-line widths, $\gamma_V^k(p, T)$, for each of the three central wings (frequency compartments) of the CO_2 $15 \mu\text{m}$ band. Likewise, we also compute the averaged absorption coefficients, κ^k , and line distances, δ^k (see Table 1).

We now assume that each wing can be approximated by the Elsasser band model (e.g., Thomas & Stamnes, 2002, Chapter 10.3):

$$\kappa(\tilde{\nu}, \delta^k, y^k) = \kappa^k f_E(\tilde{\nu}, \delta^k, y^k), \quad f_E = \frac{\sinh(2\pi y^k)}{\cosh(2\pi y^k) - \cos(2\pi \tilde{\nu} / \delta^k)}. \quad (\text{F20})$$

Here, $f_E(\tilde{\nu}, \delta^k, y^k)$ is the Elsasser form function, $y^k = \gamma_V^k / \delta^k$ is the grayness parameter, and $\tilde{\nu}$ is the frequency in units of cm^{-1} ($\tilde{\nu} = \nu / (100 c)$). This simple band model allows us to perform a pseudo line-by-line integration of Equations F10 and F11. To this end we discretize the frequency interval $[\tilde{\nu}_0^k, \dots, \tilde{\nu}_0^k + \delta^k / 2]$, where $\tilde{\nu}_0^k$ can be chosen such that $\cos(2\pi \tilde{\nu}_0^k / \delta^k) = 1$. We define $x = 2\pi \tilde{\nu} / \delta^k - 2\pi \tilde{\nu}_0^k / \delta^k$ and solve the radiative transfer equations for discrete values of x_i between $x = 0$ and $x = \pi$, multiplied by the corresponding Δx_i . The results from these spectral elements are then added and multiplied by $2N^k$, where N^k is the number of lines in the wing. This way we obtain the total energy flux density for each wing $k = 4..6$. We discretize the interval $x = 0.. \pi$ by 20 frequencies on an irregular grid, with Δx_i being smallest near $x = 0$ to ensure that the limit of sharp lines (small grayness parameter) is well resolved. Summarizing, the discrete radiative transfer equations that are solved for each spectral element $i = 1..20$ in each wing $k = 4..6$ are

$$\frac{\partial U_i^k}{\partial p} = \frac{7 - \omega^k}{g} (\kappa_i^k + \kappa^c) U_i^k + \frac{1 - \omega^k}{g} (\kappa_i^k + \kappa^c) D_i^k - \frac{2 - 2\omega^k}{g} (\kappa_i^k + \kappa^c) \pi B_i^k \quad (\text{F21})$$

$$\frac{\partial D_i^k}{\partial p} = -\frac{7 - \omega^k}{g} (\kappa_i^k + \kappa^c) D_i^k - \frac{1 - \omega^k}{g} (\kappa_i^k + \kappa^c) U_i^k + \frac{2 - 2\omega^k}{g} (\kappa_i^k + \kappa^c) \pi B_i^k. \quad (\text{F22})$$

which is analogous to Equations F14 and F15. Here,

$$B_i^k = \frac{\Delta x_i^k / \pi}{2 N^k} B^k \Leftrightarrow B^k = 2 N^k \sum_i B_i^k, \quad (\text{F23})$$

and the extinction coefficients for each of the spectral elements are

Table F1

Parameters Used in the Longwave Radiation Scheme: Frequency-Averaged Mass Extinction Coefficients, Frequency Ranges of the Bands, and Parameters to Calculate the Averaged Grayness Parameters and Single Scattering Albedos

k	κ^k ($\text{m}^2 \text{kg}^{-1}$)	$\tilde{\nu}_1^k$ (cm^{-1})	$\tilde{\nu}_2^k$ (cm^{-1})	n^k	N^k	γ_{air}^k (cm^{-1})	$\gamma_{\text{CO}_2}^k$ (cm^{-1})	δ^k (cm^{-1})	A_{12}^k (s^{-1})	χ_{air}^k (s^{-1})	χ_{O}^k (s^{-1})
1	90	980	1,100						2.73	1,500	0
2	0.08	1,300	2,100						4.66	20,000	0
3	6.0	1	540						3.12	20,000	0
4	52.14	620	667	0.7476	37	0.0727	0.0934	1.269	0.59	3,500	8×10^6
5	1270	667	671	0.7442	29	0.0730	0.0944	0.1369	1.45	7,000	1.6×10^6
6	71.75	671	715	0.7463	35	0.0697	0.0857	1.2567	1.14	7,000	1.6×10^6

$$\kappa_i^k = \kappa^k \frac{\sinh(2\pi y^k)}{\cosh(2\pi y^k) - \cos(2\pi x_i)}, \quad y^k = \gamma_V^k / \delta_k. \quad (\text{F24})$$

The net radiative energy flux densities for each compartment are

$$U^k = 2 N^k \sum_{i=1}^{20} U_i^k, \quad D^k = 2 N^k \sum_{i=1}^{20} D_i^k. \quad (\text{F25})$$

Incorporation of non-LTE effects in longwave radiative transfer is essential for CO₂ and other minor constituents in the mesosphere and thermosphere. An overview over the most sophisticated methods can be found in Feofilov and Kutepov (2012). Technical feasibility of GCMs, on the other hand, requires more approximate methods (e.g., Fomichev et al., 1998). For the mechanistic approach of the HIAMCM we use the single scattering albedos for each frequency band to describe the effects of non-LTE in the radiative transfer equation. As shown in the text of Thomas and Stamnes (2002, Chapter 4.4), this approach can be derived from microscopic theory for the idealized case of a two-level atom. According to this picture, the scattering albedo for each band can be defined as

$$\varpi^k = 1 - \frac{C_{21}^k}{C_{21}^k + A_{21} \left(1 - \exp\left(-\frac{h\nu^k}{k_B T}\right) \right)}, \quad (\text{F26})$$

where ν^k is the central frequency of the band, A_{21}^k is the frequency-averaged Einstein coefficient for spontaneous emission (calculated from the HITRAN data base, see Table F1), and C_{21}^k are the rates of inelastic collisions. The latter are defined as (e.g., Knöpfel & Becker, 2011)

$$C_{21}^k = K_{12}^k \frac{g_1^k}{g_2^k} \exp\left(\frac{h\nu^k}{k_B T}\right) [M]. \quad (\text{F27})$$

Here, K_{12}^k are quenching coefficients, g_1^k and g_2^k are the statistical weights of the two states, and $[M]$ is the number density of the molecules giving rise to inelastic collision. Usually, $[M]$ is the number density of air molecules. Atomic oxygen gives rise to additional inelastic collisions for CO₂. Using

$$K_{12}^k \propto \left(\frac{T}{T_0}\right)^{1/2} \exp\left(-\frac{h\nu^k}{K_B T}\right), \quad (\text{F28})$$

and noting that $g_1^k/g_2^k \approx 1$ up to an error of less than 5%, the rate of inelastic collisions can be written as

$$C_{21}^k = (\chi_{\text{air}}^k q_{\text{air}} + \chi_{\text{O}}^k q_{\text{O}}) \frac{\rho}{\rho_0} \left(\frac{T}{T_0}\right)^{1/2}. \quad (\text{F29})$$

Here, $\rho_0 = R_1 T_0 / p_{00}$, χ_{air}^k , and χ_{O}^k are tuning parameters (see Table F1) and $q_{\text{air}} = (R_2 - R(p)) / (R_2 - R_1)$ and $q_{\text{O}} = (R(p) - R_1) / (R_2 - R_1)$ such that $q_{\text{air}} + q_{\text{O}} = 1$.

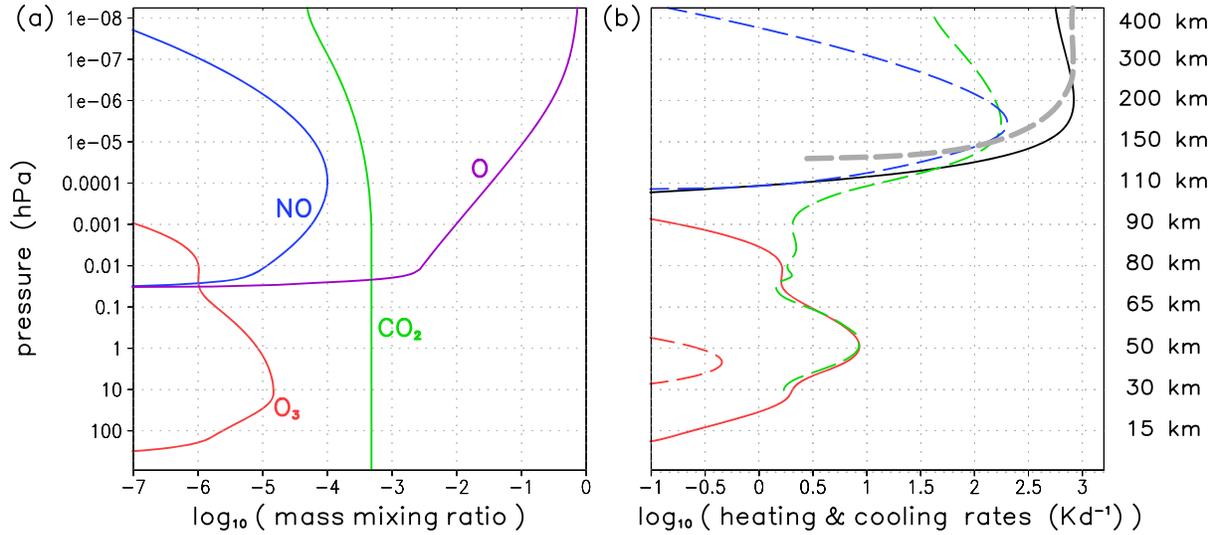


Figure F1. (a) Prescribed global-mean mass mixing ratios for ozone (red), nitric oxide (green), atomic oxygen (purple), and carbon dioxide (green). (b) Global-mean shortwave radiative heating rates (solid lines) in the middle atmosphere due to ozone (red) and in the thermosphere (black), as well as global-mean radiative cooling rates (dashed lines) due to carbon dioxide (green), ozone (red), and nitric oxide (blue). The gray dashed line shows the global-mean cooling rate due to molecular heat conduction. All heating and cooling rates are averaged over January and then plotted using logarithmic scaling.

The additional radiative transfer equations for the broadband regime are analogous to Equations F14 and F15. The broadband regime is applied to the frequency domain that is not covered by the other six frequency bands:

$$\frac{\partial U^c}{\partial p} = \frac{7/4}{g} \kappa^c U^c + \frac{1/4}{g} \kappa^c D^c - \frac{2}{g} \kappa^c \left(\sigma T^4 - \pi \sum_{k=1}^6 B^k \right) \quad (\text{F30})$$

$$\frac{\partial D^c}{\partial p} = -\frac{7/4}{g} \kappa^c D^c - \frac{1/4}{g} \kappa^c U^c + \frac{2}{g} \kappa^c \left(\sigma T^4 - \pi \sum_{k=1}^6 B^k \right). \quad (\text{F31})$$

Equations F14 and F15, F21 and F22, and F30 and F31 are solved iteratively using implicit forward Euler stepping on the staggered vertical grid of the GCM. All fluxes are computed on the half levels, and all extinction coefficients and Planck function terms are specified on full levels. We apply the usual boundary conditions: At $p = 0$ we have $D^k = 0$ for $k = 1 - 3$, $D_i^k = 0$ for $i = 1 - 20$ and $k = 4 - 6$, and $D^c = 0$. At the surface ($p = p_s$, $T = T_s$) we set $U^k = \pi B^k(T_s)$ for $k = 1 - 3$, $U_i^k = \pi B_i^k(T_s)$ for $i = 1 - 20$ and $k = 4 - 6$, and $U^c = \sigma T_s^4 - \pi \sum_{k=1}^6 B^k(T_s)$. The (predominantly negative) heating rates from the longwave radiative transfer scheme are

$$Q_{LW} = g \partial_p (D - U) = \frac{g}{\partial_p} \partial_\eta (D - U), \quad U = \sum_{k=1}^6 U^k + U^c, \quad D = \sum_{k=1}^6 D^k + D^c. \quad (\text{F32})$$

F3. Global-Mean Profiles

The performance of the radiation scheme in terms of global-mean profiles during January computed from the January control simulation is illustrated in Figures F1 and F2. Figure F1a shows the mass mixing ratios prescribed in the HIAMCM. Except for ozone (red curve), which also depends on latitude, the prescribed mass mixing ratios depend only on the vertical coordinate. The resulting shortwave and longwave radiative heating rates are shown in Figure F1b (longwave heating rates are plotted with negative sign for convenience). The heating rate from absorption by ozone (red solid line in panel b) has been tuned to match with results from comprehensive models (e.g., Fomichev et al., 2002). Likewise, ozone gives rise to some cooling in the stratopause region in the longwave regime (red dashed line). The heating from EUV absorption in the thermosphere (black solid line) has been tuned such that the altitude profile of the corresponding energy

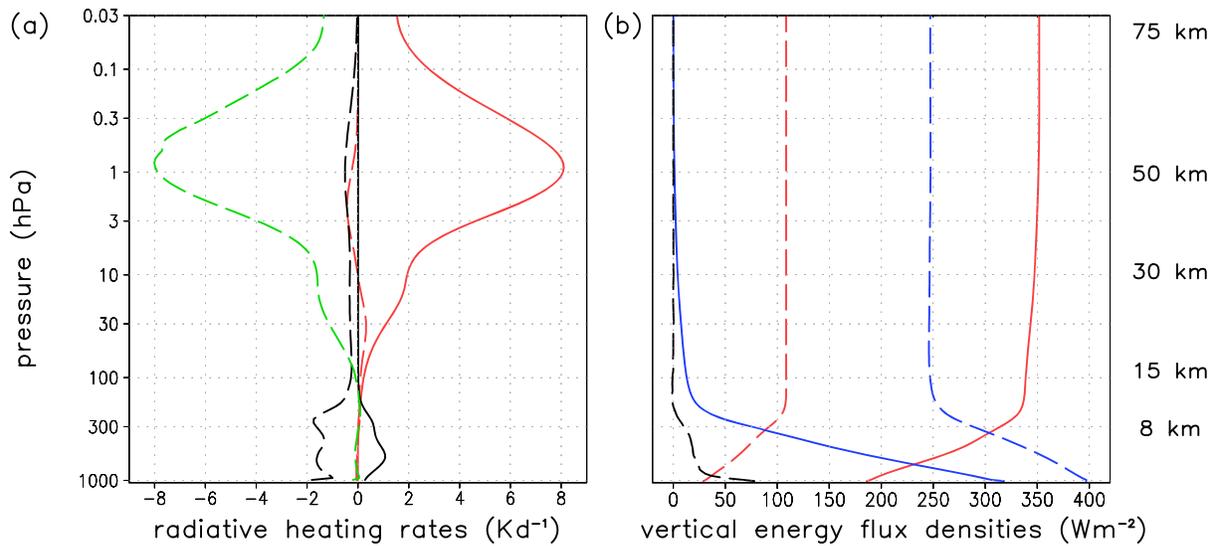


Figure F2. (a) Global-mean shortwave (solid) and longwave (dashed) radiative heating during January due to ozone (red), carbon dioxide (green), and water vapor plus clouds (black). (b) Downward (solid) and upward (dashed) vertical energy flux densities due to solar radiation (red), thermal radiation (blue), and sensible plus latent heat fluxes (black).

deposition rate (not shown) is consistent with former results of Torr et al. (1981). The longwave cooling by CO₂ (green dashed line in Figure F1b) largely balances the shortwave heating in the middle atmosphere and strongly decreases above $z \sim 150$ km (note the logarithmic scale). Additional cooling in the mesopause region and lower thermosphere is simulated by our simple incorporation of NO in the longwave transfer computations. The major cooling of the upper thermosphere is due to molecular heat conduction (plotted above $z \sim 130$ km as a gray dashed line in Figure F1b), as it should be.

Figure F2 shows the simulated global-mean diabatic heating rates (panel a) and vertical energy flux densities (panel b) due to radiative and dynamical processes from the surface to the mesosphere. The red and green curves in Figure F2a show the shortwave (solid) and longwave (dashed) heating rates due to ozone and carbon dioxide. The black lines refer to the combined shortwave (solid) and longwave (dashed) radiative heating rates from water vapor and clouds. These results are well consistent with current wisdom about the troposphere and stratosphere plus lower mesosphere. In addition to the aforementioned results for the stratosphere region, ozone gives rise to a weak longwave heating in the lower stratosphere caused by the absorption of longwave radiation emitted by the surface. Furthermore, the absorption of solar insolation by the moist phases significantly heats the troposphere. In accordance with estimates from Trenberth et al. (2009), the HIAMCM reflects about 110 W m^{-2} , where $\sim 25 \text{ W m}^{-2}$ are reflected at the surface. Likewise, the surface receives $\sim 185 \text{ W m}^{-2}$ and absorbs $\sim 160 \text{ W m}^{-2}$. Also, the greenhouse effect in the HIAMCM is simulated quite reasonably: The surface emits $\sim 400 \text{ W m}^{-2}$ in the longwave regime and absorbs $\sim 320 \text{ W m}^{-2}$. Furthermore, the surface emits $\sim 80 \text{ W m}^{-2}$ as latent and sensible heat (black dashed curve in Figure F2b). These numbers demonstrate that the HIAMCM simulates realistic global-mean energetics.

Data Availability Statement

Additional information can be found in Becker (2009), Schlutow et al. (2014), and Becker and Vadas (2018). The model data shown in this paper are available via NWRA's website (<https://www.cora.nwra.com/~erich.becker/BeckerVadas-JGRSP-2020-files>).

References

- Alexander, S. P., Sato, K., Watanabe, S., Kawatani, Y., & Murphy, D. J. (2016). Southern Hemisphere extratropical gravity wave sources and intermittency revealed by a middle-atmosphere general circulation model. *Journal of the Atmospheric Sciences*, 73, 1335–1349. <https://doi.org/10.1175/JAS-D-15-0149.1>
- Alexander, M. J., & Teitelbaum, H. (2011). Three-dimensional properties of Andes mountain waves observed by satellite: A case study. *Journal of Geophysical Research*, 116, D23110. <https://doi.org/10.1029/2011JD016151>

Acknowledgments

Model simulations were performed by the authors. A comprehensive model description is given in section 22 and in the appendices of this paper. E. B. was supported by the Leibniz Institute of Atmospheric Physics at the University of Rostock (IAP), by the Department of Earth and Planetary Science at the University of Tokyo, and by NASA Grant 80NSSC19K0834. S. L. V. was supported by NSF Grants AGS-1832988 and AGS-1552315 and by NASA Grant 80NSSC19K0836. We thank Kai Budde for evaluating the HITRAN data. Discussions with Richard Walterscheid, Joe Huba, Douglas Drob, and Aaron Riddley are gratefully acknowledged. We thank two anonymous reviewers for numerous helpful comments on the manuscript.

- Becker, E. (2003). Frictional heating in global climate models. *Monthly Weather Review*, *131*, 508–520.
- Becker, E. (2004). Direct heating rates associated with gravity wave saturation. *Journal of Atmospheric and Solar-Terrestrial Physics*, *66*, 683–696. <https://doi.org/10.1016/j.jastp.2004.01.019>
- Becker, E. (2009). Sensitivity of the upper mesosphere to the Lorenz energy cycle of the troposphere. *Journal of the Atmospheric Sciences*, *66*, 648–666. <https://doi.org/10.1175/2008JAS2735.1>
- Becker, E. (2017). Mean-flow effects of thermal tides in the mesosphere and lower thermosphere. *Journal of the Atmospheric Sciences*, *74*, 2043–2063. <https://doi.org/10.1175/JAS-D-16-0194.1>
- Becker, E., & Burkhardt, U. (2007). Nonlinear horizontal diffusion for GCMs. *Monthly Weather Review*, *135*, 1439–1454.
- Becker, E., Knöpfel, R., & Lübken, F.-J. (2015). Dynamically induced hemispheric differences in the seasonal cycle of the summer polar mesopause. *Journal of Atmospheric and Solar-Terrestrial Physics*, *129*, 128–141. <https://doi.org/10.1016/j.jastp.2015.04.014>
- Becker, E., & Vadas, S. L. (2018). Secondary gravity waves in the winter mesosphere: Results from a high-resolution global circulation model. *Journal of Geophysical Research: Atmospheres*, *123*, 2605–2627. <https://doi.org/10.1002/2017JD027460>
- Bossert, K., Fritts, D. C., Williams, B. P., Kruse, C. G., Healea, C. J., Pautet, P.-D., et al. (2017). Secondary gravity wave generation over New Zealand during the DEEPWAVE campaign. *Journal of Geophysical Research: Atmospheres*, *122*, 7834–7850. <https://doi.org/10.1002/2016JD026079>
- Brune, S., & Becker, E. (2013). Indications of stratified turbulence in a mechanistic GCM. *Journal of the Atmospheric Sciences*, *70*, 231–247. <https://doi.org/10.1175/JAS-D-12-025.1>
- Chen, C., & Chu, X. (2017). Two-dimensional Morlet wavelet transform and its application to wave recognition methodology of automatically extracting two-dimensional wave packets from lidar observations in Antarctica. *Journal of Atmospheric and Solar-Terrestrial Physics*, *162*, 28–47. <https://doi.org/10.1016/j.jastp.2016.10.016>
- Chen, C., Chu, X., McDonald, A. J., Vadas, S. L., Yu, Z., Fong, W., & Lu, X. (2013). Inertia-gravity waves in antarctica: A case study using simultaneous lidar and radar measurements at McMurdo/Scott Base (77.8°S, 166.7°E). *Journal of Geophysical Research: Atmospheres*, *118*, 2794–2808. <https://doi.org/10.1002/jgrd.50318>
- Chen, C., Chu, X., Zhao, J., Roberts, B. R., Yu, Z., Fong, W., et al. (2016). Lidar observations of persistent gravity waves with periods of 3–10 h in the Antarctic middle and upper atmosphere at McMurdo (77.83°S, 166.67°E). *Journal of Geophysical Research: Space Physics*, *121*, 1483–1502. <https://doi.org/10.1002/2015JA022127>
- Crowley, G., Jones, T. B., & Dudeney, J. R. (1987). Comparison of short period TID morphologies in Antarctica during geomagnetically quiet and active intervals. *Journal of Atmospheric and Solar-Terrestrial Physics*, *49*, 1155–1162. [https://doi.org/10.1016/0021-9169\(87\)90098-5](https://doi.org/10.1016/0021-9169(87)90098-5)
- Crowley, G., & Rodrigues, F. S. (2012). Characteristics of traveling ionospheric disturbances observed by the TIDDBIT sounder. *Radio Science*, *47*, RS0L22. <https://doi.org/10.1029/2011RS004959>
- Dalgarno, A., & Smith, F. J. (1962). The thermal conductivity and viscosity of atomic oxygen. *Planetary and Space Science*, *9*, 1–2. [https://doi.org/10.1016/0032-0633\(62\)90064-8](https://doi.org/10.1016/0032-0633(62)90064-8)
- de Wit, R. J., Janches, D., Fritts, D. C., Stockwell, R., & Coy, L. (2017). Unexpected climatological behavior of MLT gravity wave momentum flux in the lee of the Southern Andes hotspot. *Geophysical Research Letters*, *44*, 1182–1191. <https://doi.org/10.1002/2016GL072311>
- Del Genio, A. D., Schubert, G., & Straus, J. M. (1979). Gravity wave propagation in a diffusively separated atmosphere with height-dependent collision frequencies. *Journal of Geophysical Research*, *84*, 4371–4378. <https://doi.org/10.1029/JA084iA08p04371>
- Drob, D. P., Emmert, J. T., Meriwether, J. W., Makela, J. J., Doornbos, E., Conde, M., et al. (2015). An update to the Horizontal Wind Model (HWM): The quiet time thermosphere. *Earth and Space Science*, *2*, 301–319. <https://doi.org/10.1002/2014EA000089>
- Eckermann, S. D., & Preusse, P. (1999). Global measurements of stratospheric mountain waves from space. *Science*, *286*, 1534–1536.
- Ern, M., Trinh, Q. T., Kaufmann, M., Krisch, I., Preusse, P., Ungermann, J., et al. (2016). Satellite observations of middle atmosphere gravity wave absolute momentum flux and of its vertical gradient during recent stratospheric warmings. *Atmospheric Chemistry and Physics*, *16*, 9983–10,019. <https://doi.org/10.5194/acp-16-9983-2016>
- Ern, M., Trinh, Q. T., Preusse, P., Gille, J. C., Mlynarczyk, M. G., Russell, J. M. III., & Riese, M. (2018). GRACILE: A comprehensive climatology of atmospheric gravity wave parameters based on satellite limb soundings. *Earth System Science Data*, *10*, 857–892. <https://doi.org/10.5194/essd-10-857-2018>
- Feofilov, A. G., & Kutepov, A. A. (2012). Infrared radiation in the mesosphere and lower thermosphere: Energetic effects and remote sensing. *Surveys in Geophysics*, *33*(6), 1231–1280.
- Fleming, E. L., Chandra, S., Barnett, J. J., & Corney, M. (1990). Zonal mean temperature, pressure, zonal wind and geopotential height as functions of latitude. *Advances in Space Research*, *10*(12), 11–59. [https://doi.org/10.1016/0273-1177\(90\)90386-E](https://doi.org/10.1016/0273-1177(90)90386-E)
- Fomichev, V. I., Blanchet, J.-P., & Turner, D. S. (1998). Matrix parameterization of the 15 μm CO₂ band cooling in the middle and upper atmosphere for variable CO₂ concentration. *Journal of Geophysical Research*, *103*(11), 11,505–11,528.
- Fomichev, V. I., Ward, W. E., Beagley, S. R., McLandress, C., McConnell, J. C., McFarlane, N. A., & Shepherd, T. G. (2002). Extended Canadian Middle Atmosphere Model: Zonal-mean climatology and physical parameterizations. *Journal of Geophysical Research*, *107*(D10), ACL 9-1–ACL 9-14. <https://doi.org/10.1029/2001JD000479>
- Forbes, J. M., Bruinsma, S. L., Doornbos, E., & Zhang, X. (2016). Gravity wave-induced variability of the middle thermosphere. *Journal of Geophysical Research: Space Physics*, *121*, 6914–6932. <https://doi.org/10.1002/2016JA022923>
- Frissell, N. A., Baker, J. B. H., Ruohoniemi, J. M., Greenwald1, R. A., Gerrard2, A. J., Miller, E. S., & West, M. L. (2016). Sources and characteristics of medium-scale traveling ionospheric disturbances observed by high-frequency radars in the north american sector. *Journal of Geophysical Research: Space Physics*, *121*, 3722–3739. <https://doi.org/10.1002/2015JA022168>
- Fritts, D. C., & Alexander, M. J. (2003). Gravity wave dynamics and effects in the middle atmosphere. *Reviews of Geophysics*, *41*, 1003. <https://doi.org/10.1029/2001/RG000106>
- Fritts, D. C., Dong, W., Lund, T. S., Wieland, S., & Laughman, B. (2020). Self-acceleration and instability of gravity wave packets: 3. Three-dimensional packet propagation, secondary gravity waves, momentum transport, and transient mean forcing in tidal winds. *Journal of Geophysical Research: Atmospheres*, *125*, e2019JD030692. <https://doi.org/10.1029/2019JD030692>
- Fritts, D. C., Smith, R. B., Taylor, M. J., Doyle, J. D., Eckermann, S. D., Dörnbrack, A., et al. (2016). The Deep Propagating Gravity Wave Experiment (DEEPWAVE): An airborne and ground-based exploration of gravity wave propagation and effects from their sources throughout the lower and middle atmosphere. *Bulletin of the American Meteorological Society*, *97*(3), 425–453. <https://doi.org/10.1175/BAMS-D-14-00269.1>
- Fritts, D. C., & Vadas, S. L. (2008). Gravity wave penetration into the thermosphere: Sensitivity to solar cycle variations and mean winds. *Annales Geophysicae*, *26*, 3841–3861.

- Fritts, D. C., Vadas, S. L., Wan, K., & Werne, J. A. (2006). Mean and variable forcing of the middle atmosphere by gravity waves. *Journal of Atmospheric and Solar-Terrestrial Physics*, *68*, 247–265. <https://doi.org/10.1016/j.jastp.2005.04.010>
- Gassmann, A. (2018). Entropy production due to subgrid-scale thermal fluxes with application to breaking gravity waves. *Quarterly Journal of the Royal Meteorological Society*, *144*, 499–510. <https://doi.org/10.1002/qj.3221>
- Gossard, E. E., & Hooke, W. H. (1975). *Waves in the atmosphere* (p. 112). New York: Elsevier.
- Heale, C. J., Bossert, K., Vadas, S. L., Hoffmann, L., Dörnbrack, A., Stober, G., et al. (2020). Secondary gravity waves generated by breaking mountain waves over Europe. *Journal of Geophysical Research: Atmospheres*, *125*, e2019JD031662. <https://doi.org/10.1029/2019JD031662>
- Heale, C. J., Walterscheid, R. L., & Snively, J. B. (2018). Localization effects on the dissipation of gravity wave packets in the upper mesosphere and lower thermosphere. *Journal of Geophysical Research: Atmospheres*, *123*, 8915–8935. <https://doi.org/10.1029/2017JD027617>
- Hedin, A. E. (1991). Extension of the MSIS thermosphere model into the middle and lower atmosphere. *Journal of Geophysical Research*, *96*(A2), 1159–1172.
- Hendricks, E. A., Doyle, J. D., Eckermann, S. D., Jiang, Q., & Reinecke, P. A. (2014). What is the source of the stratospheric gravity wave belt in austral winter. *Journal of the Atmospheric Sciences*, *71*, 1583–1592. <https://doi.org/10.1175/JAS-D-13-0332.1>
- Hocke, K., & Schlegel, K. (1996). A review of atmospheric gravity waves and travelling ionospheric disturbances: 1982–1995. *Annales Geophysicae*, *14*, 917–940.
- Holtslag, A. A. M., & Boville, B. A. (1993). Local versus non-local boundary-layer diffusion in a global climate model. *Journal of Climate*, *6*, 1825–1842.
- Hong, S.-S., & Lindzen, R. S. (1976). Solar semidiurnal tide in the thermosphere. *Journal of the Atmospheric Sciences*, *33*, 135–153.
- Howard, L. N. (1961). Note on a paper of John W. Miles. *Journal of Fluid Mechanics*, *10*, 509–512.
- Huang, K. M., Liu, A. Z., Zhang, S. D., Yi, F., Huang, C. M., Gong, Y., et al. (2017). Simultaneous upward and downward propagating inertia-gravity waves in the MLT observed at Andes Lidar Observatory. *Journal of Geophysical Research: Atmospheres*, *122*, 2812–2830. <https://doi.org/10.1002/2016JD026178>
- Janjic, Z. I., Gerrity, J. P. Jr., & Nickovic, S. (2001). An alternative approach to nonhydrostatic modeling. *Monthly Weather Review*, *129*, 1164–1178.
- Jiang, J. H., Wu, D. L., & Eckermann, S. D. (2002). Upper Atmosphere Research Satellite (UARS) MLS observation of mountain waves over the Andes. *Journal of Geophysical Research*, *107*(D20), 8273. <https://doi.org/10.1029/2002JD002091>
- Klingbeil, K., & Burchard, H. (2013). Implementation of a direct nonhydrostatic pressure gradient discretisation into a layered ocean model. *Ocean Modelling*, *65*, 64–77. <https://doi.org/10.1016/j.ocemod.2013.02.002>
- Knöpfel, R., & Becker, E. (2011). An idealized radiative transfer scheme for use in a mechanistic general circulation model from the surface up to the mesopause region. *Journal of Quantitative Spectroscopy & Radiative Transfer*, *112*, 1460–1478. <https://doi.org/10.1016/j.jqsrt.2011.02.014>
- Kuo, H. L. (1965). On formation and intensification of tropical cyclones through latent heat release by cumulus convection. *Journal of the Atmospheric Sciences*, *22*, 40–63.
- Lindzen, R. S. (1981). Turbulence and stress owing to gravity wave and tidal breakdown. *Journal of Geophysical Research*, *86*, 9707–9714.
- Liu, H.-L. (2017). Large wind shears and their implications for diffusion in regions with enhanced static stability: The mesopause and the tropopause. *Journal of Geophysical Research: Atmospheres*, *122*, 9579–9590. <https://doi.org/10.1002/2017JD026748>
- Liu, H.-L., Bardeen, C. G., Foster, B. T., Lauritzen, P., Liu, J., Lu, G., et al. (2018). Development and validation of the Whole Atmosphere Community Climate Model with thermosphere and ionosphere extension (WACCM-X 2.0). *Journal of Advances in Modeling Earth Systems*, *10*, 381–402. <https://doi.org/10.1002/2017MS001232>
- Liu, H.-L., McNerney, J. M., Santos, S., Lauritzen, P. H., Taylor, M. A., & Pedatella, N. M. (2014). Gravity waves simulated by high-resolution Whole Atmosphere Community Climate Model. *Geophysical Research Letters*, *41*, 9106–9112. <https://doi.org/10.1002/2014GL062468>
- Liu, X., Xu, J., Yue, J., Vadas, S. L., & Becker, E. (2019). Orographic primary and secondary gravity waves in the middle atmosphere from 16-year SABER observations. *Geophysical Research Letters*, *46*, 4512–4522. <https://doi.org/10.1029/2019GL082256>
- Liu, C. H., & Yeh, K. C. (1969). Effect of ion drag on propagation of acoustic-gravity waves in the atmospheric F region. *Journal of Geophysical Research*, *74*(2), 2248–2255.
- Lübken, F.-J. (1997). Seasonal variation of turbulent energy dissipation rates at high latitudes as determined by in situ measurements of neutral density fluctuations. *Journal of Geophysical Research*, *102*, 13,441–13,456.
- Lund, T. S., & Fritts, D. C. (2012). Numerical simulation of gravity wave breaking in the lower thermosphere. *Journal of Geophysical Research*, *117*, D21105. <https://doi.org/10.1029/2012JD017536>
- McLandress, C., Ward, W. E., Fomichev, V. I., Semeniuk, K., Beagley, S. R., McFarlane, N. A., & Shepherd, T. G. (2006). Large-scale dynamics of the mesosphere and lower thermosphere: An analysis using the extended Canadian Middle Atmosphere Model. *Journal of Geophysical Research*, *111*, D17111. <https://doi.org/10.1029/2005JD006776>
- Miles, J. W. (1961). On the stability of heterogeneous shear flows. *Journal of Fluid Mechanics*, *10*, 496–508.
- Miyoshi, Y., Fujiwara, H., Jin, H., & Shinagawa, H. (2014). A global view of gravity waves in the thermosphere simulated by a general circulation model. *Journal of Geophysical Research: Space Physics*, *119*, 5807–5820. <https://doi.org/10.1002/2014JA019848>
- Muraschko, J., Fruman, M. D., Achatz, U., Hickel, S., & Toledoc, Y. (2015). On the application of Wenzel-Kramer-Brillouin theory for the simulation of the weakly nonlinear dynamics of gravity waves. *Quarterly Journal of the Royal Meteorological Society*, *141*, 676–697. <https://doi.org/10.1002/qj.2381>
- Nicolls, M. J., Vadas, S. L., Aponte, N., & Sulzer, M. P. (2014). Horizontal wave parameters of daytime thermospheric gravity waves and E region neutral winds over Puerto Rico. *Journal of Geophysical Research: Space Physics*, *119*, 575–600. <https://doi.org/10.1002/2013JA018988>
- O'Sullivan, D., & Dunkerton, T. J. (1995). Generation of inertia-gravity waves in a simulated life-cycle of baroclinic instability. *Journal of the Atmospheric Sciences*, *52*, 3695–3716.
- Olbers, D., Eden, C., Becker, E., Pollmann, F., & Jungclaus, J. (2019). The IDEMIX Model: Parameterization of internal gravity waves for circulation models of ocean and atmosphere. In C. Eden & A. Iske (Eds.), *Energy Transfers in Atmosphere and Ocean, Mathematics of Planet Earth* (Vol. 1, pp. 87–125). Cham: Springer. https://doi.org/10.1007/978-3-030-05704-6_3
- Park, J., Lühr, H., Lee, C., Kim, Y. H., Jee, G., & Kim, J.-H. (2014). A climatology of medium-scale gravity wave activity in the midlatitude/low-latitude daytime upper thermosphere as observed by CHAMP. *Journal of Geophysical Research: Space Physics*, *19*, 2187–2196. <https://doi.org/10.1002/2013JA019705>

- Pierrehumbert, R. T. (2011). Infrared radiation and planetary temperature. *Physics Today*, *64*(1), 33–38. <https://doi.org/10.1063/1.3541943>
- Plougonven, R., & Zhang, F. (2014). Internal gravity waves from atmospheric jets and fronts. *Reviews of Geophysics*, *52*, 33–76. <https://doi.org/10.1002/2012RG000419>
- Roble, R. G., & Ridley, E. C. (1994). A thermosphere-ionosphere-mesosphere-electrodynamics general circulation model (time-GCM): Equinox solar cycle minimum simulations (30–500 km). *Geophysical Research Letters*, *21*, 417–420.
- Rothman, L. S., Gordon, I. E., Babikov, Y., Barbe, A., Benner, D. C., Bernath, P. F., et al. (2013). The HITRAN2012, molecular spectroscopic database. *Journal of Quantitative Spectroscopy & Radiative Transfer*, *130*, 4–50. <https://doi.org/10.1016/j.jqsrt.2013.07.002>
- Sato, K., Tatenno, S., Watanabe, S., & Kawatani, Y. (2012). Gravity wave characteristics in the Southern Hemisphere revealed by a high-resolution middle-atmosphere general circulation model. *Journal of the Atmospheric Sciences*, *69*, 1378–1396. <https://doi.org/10.1175/JAS-D-11-0101.1>
- Sato, K., & Yoshiki, M. (2008). Gravity wave generation around the polar vortex in the stratosphere revealed by 3-hourly radiosonde observations at Syowa Station. *Journal of the Atmospheric Sciences*, *65*, 3719–3735. <https://doi.org/10.1175/2008JAS2539.1>
- Schaefer-Rolffs, U., & Becker, E. (2018). Scale-invariant formulation of momentum diffusion for high-resolution atmospheric circulation models. *Monthly Weather Review*, *146*, 1045–1062. <https://doi.org/10.1175/MWR-D-17-0216.1>
- Schlutow, M., Becker, E., & Körnich, H. (2014). Positive definite and mass conserving tracer transport in spectral GCMs. *Journal of Geophysical Research: Atmospheres*, *119*, 11,562–11,577. <https://doi.org/10.1002/2014JD021661>
- Shibuya, R., Sato, K., Tsutsumi, M., Sato, T., Tomikawa, Y., Nishimura, K., & Kohma, M. (2017). Quasi-12 h inertia-gravity waves in the lower mesosphere observed by the pansy radar at Syowa Station (39.6°E, 69.0°S). *Atmospheric Chemistry and Physics*, *17*, 6455–6476. <https://doi.org/10.5194/acp-17-6455-2017>
- Simmons, A. J., & Burridge, D. M. (1981). An energy and angular momentum conserving vertical finite-difference scheme and hybrid vertical coordinates. *Monthly Weather Review*, *109*, 758–766.
- Smagorinsky, J. (1993). Some historical remarks on the use of nonlinear viscosities. In B. Galperin, & Orszag, A. St. (Eds.), *Large eddy simulation of complex engineering and geophysical flows* (pp. 3–36). Cambridge: Cambridge University Press.
- Smith, A. K. (2012). Global dynamics of the MLT. *Surveys Geophysics*, *33*, 1177–1230. <https://doi.org/10.1007/s10712-012-9196-9>
- Thomas, G. E., & Stamnes, K. (2002). *Radiative Transfer in the Atmosphere and Ocean, Cambridge Atmospheric and Space Science Series* (p. 517). Cambridge: Cambridge University Press.
- Torr, M. R., Richards, P. G., & Torr, D. G. (1981). Solar EUV energy budget of the thermosphere. *Advances in Space Research*, *1*, 53–61. [https://doi.org/10.1016/0273-1177\(81\)90417-8](https://doi.org/10.1016/0273-1177(81)90417-8)
- Trenberth, K. E., Fasullo, J. T., & Kiehl, J. (2009). Earth's global energy budget. *Bulletin of the American Meteorological Society*, *90*, 311–323. <https://doi.org/10.1175/2008BAMS2634.1>
- Trinh, Q. T., Ern, M., Doornbos, E., Preusse, P., & Riese, M. (2018). Satellite observations of middle atmosphere-thermosphere vertical coupling by gravity waves. *Annales Geophysicae*, *36*, 425–444. <https://doi.org/10.5194/angeo-36-425-2018>
- Triplett, C. C., Collins, R. L., Nielsen, K., Harvey, V. L., & Mizutani, K. (2017). Role of wind filtering and unbalanced flow generation in middle atmosphere gravity wave activity at Chatanika Alaska. *Atmosphere*, *8*(2), 27. <https://doi.org/10.3390/atmos8020027>
- Vadas, S. L. (2007). Horizontal and vertical propagation and dissipation of gravity waves in the thermosphere from lower atmospheric and thermospheric sources. *Journal of Geophysical Research*, *112*, A06305. <https://doi.org/10.1029/2006JA011845>
- Vadas, S. L., & Becker, E. (2018). Numerical modeling of the excitation, propagation, and dissipation of primary and secondary gravity waves during wintertime at McMurdo station in the Antarctic. *Journal of Geophysical Research: Atmospheres*, *123*, 9326–9369. <https://doi.org/10.1029/2017JD027974>
- Vadas, S. L., & Becker, E. (2019). Numerical modeling of the generation of tertiary gravity waves in the mesosphere and thermosphere during strong mountain wave events over the Southern Andes. *Journal of Geophysical Research: Space Phys.*, *124*, 7687–7718. <https://doi.org/10.1029/2019JA026694>
- Vadas, S. L., & Crowley, G. (2010). Sources of the traveling ionospheric disturbances observed by the ionospheric TIDDBIT sounder near Wallops Island on October 30, 2007. *Journal of Geophysical Research*, *115*, A07324. <https://doi.org/10.1029/2009JA015053>
- Vadas, S. L., & Crowley, G. (2017). Neutral wind and density perturbations in the thermosphere created by gravity waves observed by the TIDDBIT sounder. *Journal of Geophysical Research: Space Physics*, *122*, 6652–6678. <https://doi.org/10.1002/2016JA023828>
- Vadas, S. L., Fritts, D. C., & Alexander, M. J. (2003). Mechanisms for the generation of secondary waves in wave breaking regions. *Journal of the Atmospheric Sciences*, *60*, 194–214. <https://doi.org/10.1029/2004JD005574>
- Vadas, S. L., & Liu, H.-L. (2009). Generation of large-scale gravity waves and neutral winds in the thermosphere from the dissipation of convectively generated gravity waves. *Journal of Geophysical Research*, *114*, A10310. <https://doi.org/10.1029/2009JA014108>
- Vadas, S. L., & Liu, H.-L. (2013). Numerical modeling of the large-scale neutral and plasma responses to the body forces created by the dissipation of gravity waves from 6 h of deep convection in Brazil. *Journal of Geophysical Research*, *118*, 2593–2617. <https://doi.org/10.1002/jgra.50249>
- Vadas, S. L., Liu, H.-L., & Lieberman, R. S. (2014). Numerical modeling of the global changes to the thermosphere and ionosphere from the dissipation of gravity waves from deep convection. *Journal of Geophysical Research*, *119*, 7762–7793. <https://doi.org/10.1002/2014JA020280>
- Vadas, S. L., & Nicolls, M. J. (2009). Temporal evolution of neutral, thermospheric winds and plasma response using PFISR measurements of gravity waves. *Journal of Atmospheric and Solar-Terrestrial Physics*, *71*, 740–770. <https://doi.org/10.1016/j.jastp.2009.01.011>
- Vadas, S. L., Xu, S., Yue, J., Bossert, K., Becker, E., & Baumgarten, G. (2019). Characteristics of the quiet-time hotspot gravity waves observed by GOCE over the Southern Andes on 5 July 2010. *Journal of Geophysical Research: Space Physics*, *124*, 7034–7061. <https://doi.org/10.1029/2019JA026693>
- Vadas, S. L., Zhao, J., Chu, X., & Becker, E. (2018). The excitation of secondary gravity waves from local body forces: Theory and observation. *Journal of Geophysical Research: Atmospheres*, *123*, 9296–9325. <https://doi.org/10.1029/2017JD027970>
- Vargas, F., Swanson, G., Liu, A., & Pautet, D. (2016). Evidence of the excitation of a ring-like gravity wave in the mesosphere over the Andes Lidar Observatory. *Journal of Geophysical Research: Atmospheres*, *121*, 8896–8912. <https://doi.org/10.1002/2016JD024799>
- Watanabe, S., & Miyahara, S. (2009). Quantification of the gravity wave forcing of the migrating diurnal tide in gravity wave-resolving general circulation model. *Journal of Geophysical Research*, *114*, D07110. <https://doi.org/10.1029/2008JD011218>
- Watanabe, S., Sato, K., & Takahashi, M. (2006). A general circulation model study of the orographic gravity waves over Antarctica excited by katabatic winds. *Journal of Geophysical Research*, *111*, D18104. <https://doi.org/10.1029/2005JD006851>
- Wu, D. L., & Waters, J. W. (1996). Satellite observations of atmospheric variances: A possible indication of gravity waves. *Geophysical Research Letters*, *23*(24), 3631–3634. <https://doi.org/10.1029/96GL02907>

- Yiğit, E., & Medvedev, A. S. (2017). Influence of parameterized small-scale gravity waves on the migrating diurnal tide in Earth's thermosphere. *Journal of Geophysical Research: Space Physics*, *122*, 4846–4864. <https://doi.org/10.1002/2017JA024089>
- Yiğit, E., Medvedev, A. S., Aylward, A. D., Hartogh, P., & Harris, M. J. (2009). Modeling the effects of gravity wave momentum deposition on the general circulation above the turbopause. *Journal of Geophysical Research*, *114*, D07101. <https://doi.org/10.1029/2008JD011132>
- Zhang, F. (2004). Generation of mesoscale gravity waves in upper-tropospheric jet-front systems. *Journal of the Atmospheric Sciences*, *61*, 440–457.
- Zhao, J., Chu, X., Chen, C., Lu, X., Fong, W., Yu, Z., et al. (2017). Lidar observations of stratospheric gravity waves from 2011 to 2015 at McMurdo (77.84°S, 166.69°E), Antarctica: 1. Vertical wavelengths, periods, and frequency and vertical wavenumber spectra. *Journal of Geophysical Research: Atmospheres*, *122*, 5041–5062. <https://doi.org/10.1002/2016JD026368>