Highlights

Multi-step vertical coupling via gravity waves from the lower to the upper atmosphere

Erich Becker, Sharon L. Vadas, Xinzhao Chu

- Multi-step vertical coupling (MSVC) from primary to high-order gravity waves (GWs) is crucial for understanding the observed prevailing winds in the winter mesopause region and to explain observed GWs in the mesosphere and thermosphere.

- For small-enough spatial and temporal scales, observed GW spectra can be interpreted as stratified macro turbulence resulting from wave breaking.

- Higher-order GWs in the winter thermosphere explain observed daytime traveling ionospheric disturbances during periods of low geomagnetic activity.

- MSVC in the winter middle and upper atmosphere correlates with the strength of the polar vortex.

- The disturbances in the thermosphere and ionosphere that were caused by the Hunga Tonga-Hunga Ha'apai volcanic eruption were due to medium-to-large-scale secondary GWs.

Multi-step vertical coupling via gravity waves from the lower to the upper atmosphere

Erich Becker[a], Sharon L. Vadas[a], Xinzhao Chu[b]

[a] *NorthWest Research Associates, Boulder Office, Boulder, CO, USA*
[b] *Cooperative Institute of Research in Environmental Sciences & Department of Aerospace Engineering Sciences, University of Colorado Boulder, Boulder, CO, USA*

Abstract

We review the mechanism of multi-step vertical coupling (MSVC) via secondary and higher-order gravity waves (GWs), and its relevance for observed GW perturbations and the circulation in the upper mesosphere and thermosphere. Since the momentum deposition following the breaking or dissipation of a GW packet is localized in space and time, it leads to an imbalance in the ambient flow which in turn results in the generation of secondary or higher-order GWs. This local "body-force" (LBF) mechanism is essential for MSVC. We argue that small-scale secondary GWs resulting directly from GW instability form a macro-turbulent cascade that leads to the LBF. We present a simple scale analysis supporting this interpretation with respect to observed GW spectra. Several examples of MSVC are reviewed. These include 1) an explanation of the observed persistent GWs and prevailing eastward winds in the winter mesopause region at middle to high latitudes via secondary GWs, 2) evidence that many of the daytime traveling ionospheric disturbances in the F region during winter and low geomagnetic activity are driven by higher-order GWs from MSVC, 3) the dependence of MSVC during wintertime on the strength of the polar vortex, and 4) the secondary GW disturbances in the thermosphere and ionospheric that were triggered by the Tonga volcanic eruption on 15 January 2022. Furthermore, we describe the GW-resolving whole-atmosphere model that was primarily used in corresponding studies of MSVC, and we discuss some open questions.

Keywords: Multi-step vertical coupling (MSVC), secondary and higher-order gravity waves (GWs), local body-force mechanism, universal behavior of GW spectra, observations and modeling of MSVC

Preprint submitted to Encyclopedia of Atmospheric Sciences *January 15, 2025*

Contents

1. Introduction

The global circulation in the upper mesosphere is mainly driven by the wave-mean flow interaction due to internal gravity waves (GWs) (Lindzen, 1981; Holton, 1983). Further contributions result from in-situ generated planetary waves (McLandress et al., 2006). The circulation is strongly driven by thermal tides in the lower thermosphere, and by ion drag at higher altitudes, while the average GW drag is minor in the mid and upper thermosphere (Becker, 2017; Becker et al., 2022a; Becker and Oberheide, 2023; Liu et al., 2024a). According to conventional wisdom (e.g., Smith, 2012; Becker, 2012), the GWs relevant for the circulation and variability in the stratosphere and mesosphere are of tropospheric origin. That is, they are primary GWs generated by flow over orography, deep moist convection, and spontaneous emission (fronts and jets) (see reviews of Fritts and Alexander, 2003; Plougonven and Zhang, 2014). In addition, primary GWs can be generated in the stratosphere by the polar vortex jet (e.g., Sato and Yoshiki, 2008; Sato et al., 2012; Becker et al., 2022b; Vadas et al., 2023a).

Recent studies suggest that the effects of GWs from "below" (i.e., from the troposphere and statosphere) in the winter upper mesosphere and winter thermosphere during periods of low geomagnetic activity are due to secondary and higher-order GWs, not primary GWs (e.g., Becker and Vadas, 2018;

Vadas and Becker, 2019; Becker and Vadas, 2020; Xu et al., 2021; Becker et al., 2022a; Vadas et al., 2024). According to theory (Vadas and Fritts, 2002; Vadas et al., 2003; Vadas, 2013; Vadas et al., 2018), secondary GWs are excited from the imbalances that are created by the localized (in space and time) wave-mean flow interactions (momentum and energy deposition) that result from the breakdown of primary GW packets. Even though this generation mechanism creates a broad spectrum of waves, the majority of the secondary GWs have larger scales than the primary GWs. The secondary GWs propagate upward and downward, and into all horizontal directions away from the source region, except for the direction perpendicular to the body force direction. They can account for significant non-local transport of momentum and energy if their propagation directions and phase speeds in relation to the background wind yield near-conservative upward propagation for a few density scale heights. In such a case, amplitude growth due to decreasing density with increasing height and wind shear (e.g., from tides) will eventually cause the secondary GWs to break, or to dissipate directly from molecular viscosity in the thermosphere, resulting in significant wave-mean flow interaction. When this process is sufficiently localized in space and time, tertiary GWs will be generated, and so forth. The vertical coupling that results from primary to secondary and higher-order (tertiary, etc) GWs has been dubbed "multi-step vertical coupling" (MSVC) (Vadas and Becker, 2019). The overall idea of this process is illustrated by the schematic in Fig. 1 for the winter hemisphere.

While GWs dissipate directly from kinematic molecular viscosity and heat conduction at high altitudes in the thermosphere (Vadas, 2007), the dissipation scales are much smaller in the middle atmosphere where GWs undergo complicated breaking processes when they reach a certain level of dynamic instability. Theories have been brought forward to describe this process in parametric form for application in global models (e.g. Lindzen, 1981; Hines, 1997; Medvedev and Klaassen, 2000; Becker and McLandress, 2009), and high-resolution numerical simulations performed under idealized conditions have provided an advanced understanding of the breaking processes that occur under various circumstances regarding the background flow and incident GW characteristics (e.g. Achatz, 2007c,b,a; Dong et al., 2020). Most importantly, the nonlinear interactions associated with GW breaking generate smaller-scale GWs which in turn generate even smaller-scale GWs due to nonlinear interactions and so forth. In the statistical mean, this process must be characterized by a forward energy cascade to higher wavenumbers (smaller

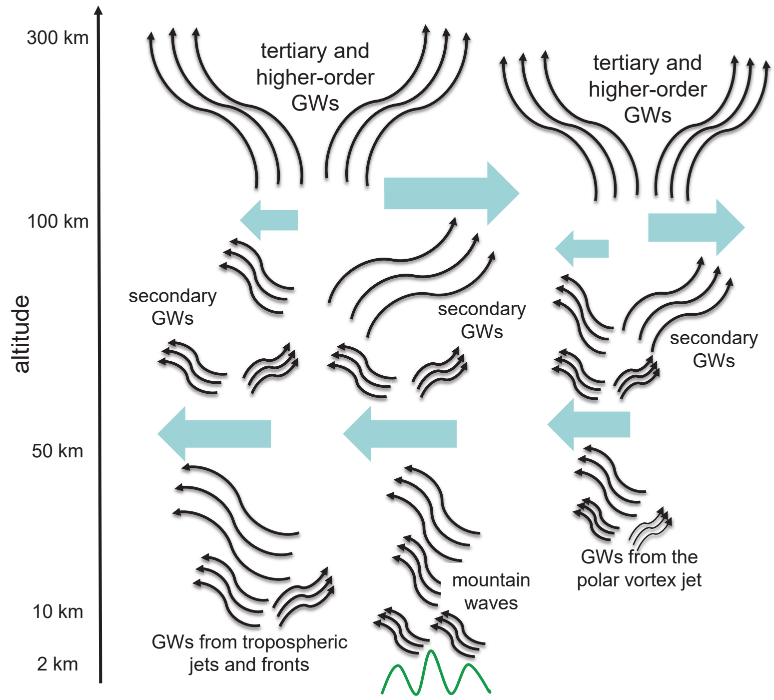

Figure 1: Schematic of multi-step vertical coupling (MSVC) in the winter hemisphere. Medium-scale primary GWs are generated by tropospheric jets and front, by flow over orography (indicated in green), and by the polar vortex jet. While mountain waves typically have westward intrinsic propagation directions, the other primary GWs can propagate into all horizontal directions. This is indicated by the black "wave-like" arrows below $z \sim 50$ km. Mainly those primary GWs having westward propagation components propagate to the upper stratosphere and mesosphere where they dissipate from dynamic instability. The resulting westward drag components are indicated by the thick blue arrows. Due to the localized and intermittent character of this momentum deposition, secondary GWs are generated. With increasing height in the mesosphere, mainly those secondary GWs that have eastward propagation components reach the mesopause region. Here, these secondary GWs dissipate from the dynamic instability induced by the vertical wind shears associated with the semi-diurnal tide and traveling planetary waves, resulting in an eastward GW drag on average with regard to the zonal direction. This process leads to the generation of tertiary GWs, which propagate to higher altitudes in the thermosphere where they dissipate and may generate other higher-order GWs. Overall, the higher-order GWs in the thermosphere propagate into all horizontal direction. Those GWs propagating against the large-scale diurnal tidal winds in the winter thermosphere have the largest amplitudes at F region altitudes (above $z \sim 250$ km). At these altitudes, the GWs dissipate directly from molecular viscosity and heat conduction. After Vadas and Becker (2019).

scales) that is initiated by the instability of the incident GW packet. These small-scale GWs must be distinguished from the secondary and higher-order GWs that give rise to MSVC. Even though some small-scale secondary GWs may also contribute to vertical coupling (Fritts et al., 2020), we will consider the small-scale GWs that contribute to the cascade to turbulence to be local. In other words, it is likely that most small-scale secondary and higher-order GWs do not transport momentum and energy far from the breaking region.

In this paper we review our current knowledge about MSVC and its role in wave phenomena and the circulation in the middle and upper atmosphere. In Sec. 2 we briefly review the aforementioned body-force mechanism. The importance of this mechanism for vertical coupling is further supported by showing that the observed universal behavior of observed GW spectra is likely a consequence of a macro-turbulent inertial range (Sec. 3). In Sec. 4 we give a brief description of the GW-resolving whole-atmosphere model that has mainly been used in published studies of MSVC. This is followed by a review of examples of MSVC in the literature (Sec. 5). We conclude with a summary and brief discussion of open questions in (Sec. 6).

2. Local body-force mechanism

The mechanism of a dissipating GW packet giving rise to secondary (or higher-order) GWs was first proposed by Vadas and Fritts (2002) and Vadas et al. (2003). In the latter paper, an idealized quasi-linear theory was presented in which the flow response to an imposed horizontal acceleration that was localized in space and time, dubbed local "body force" (LBF), was calculated using a Fourier-Laplace transform following Vadas and Fritts (2001). This analysis was revised in Vadas (2013) to include compressibility, and was applied in Vadas et al. (2018) where results from theory were compared to the first observational evidence of the LBF mechanism provided by lidar data from McMurdo Station (Antarctica). Figure 2 summarizes the major aspects of the flow response to a LBF. Several points should be noted:

(1) The solution consists of an ambient (mean) flow response plus a broad spectrum of new (secondary) GWs. The ambient flow response is characterized by two counter-rotating cells (or vortices) in the horizontal plane. This response is simply the direct result of the acceleration by the body force and mass conservation, leading to return flows that are anti-parallel to the body force (see Fig. 2a).

(2) The GWs excited by a LBF result from the imbalance that is generated

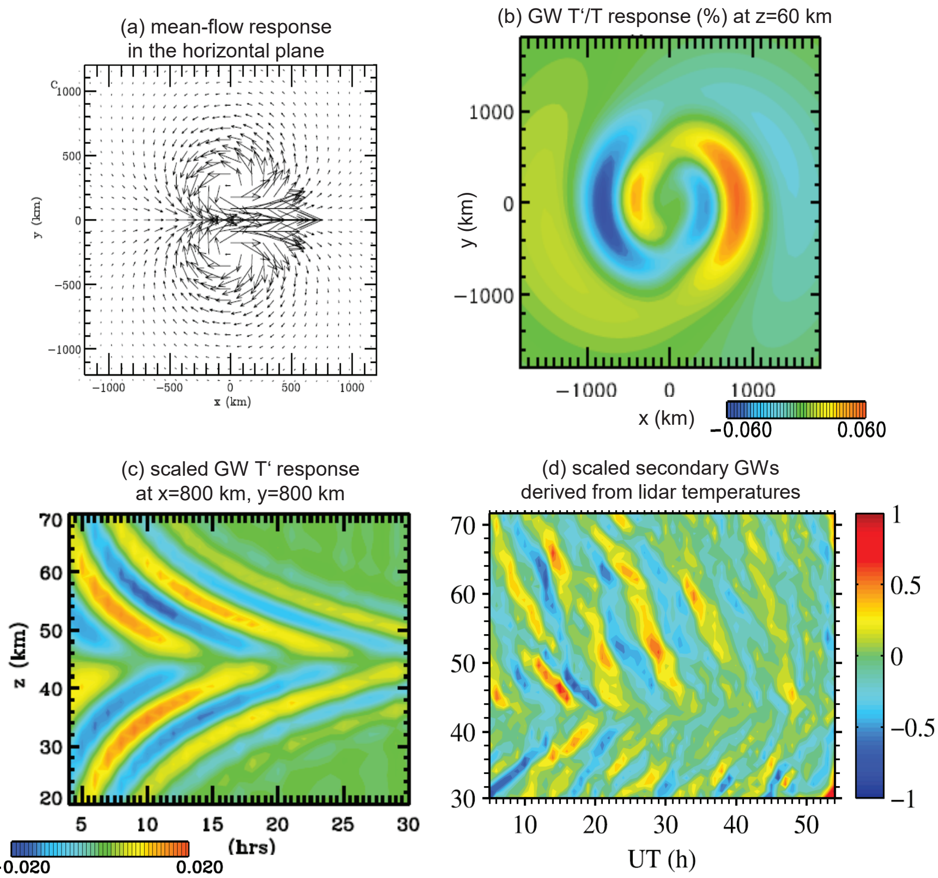

Figure 2: Illustration of the local body-force (LBF) mechanism. (a) Horizontal snapshot of the ambient horizontal-wind response predicted by theory at the height of a zonal LBF ($z = 45\,\mathrm{km}$) and about 16 hr after the force is finished. The body force has full vertical, horizontal, and temporal widths of 8 km, 800 km, and 2 hours, respectively. Its maximum acceleration is 120 $\mathrm{m\,s^{-1}d^{-1}}$. (b) Corresponding GW relative temperature response at $z = 60\,\mathrm{km}$ and $t = 4\,\mathrm{hr}$. (c) Corresponding time-height plot of the density-scaled relative GW temperature response, $\sqrt{\rho}T'/T$, in units of $\sqrt{\mathrm{g\,m^3}}$ as predicted by theory. (d) Density-scaled secondary GWs derived from lidar temperature observations at McMurdo Station (178°E, 78°S), starting on 18 June 2014. The unit is $10^{-3}\sqrt{\mathrm{kg\,m^3}}$. See Vadas et al. (2018, their Figs. 9, 13, and 15) for further details.

in the ambient flow. These secondary GWs consist of a broad spectrum of waves. When the duration of the LBF is short enough, the largest GW amplitudes occur at horizontal and vertical wavelengths that are about twice the length/width and height of the body force, respectively. The secondary GWs propagate into all directions except perpendicular to the body force (Fig. 2b).

(3) At a location that is horizontally displaced in a direction not perpendicular of the body force, the GW response appears as a “fishbone structure” in time-height plots of the temperature, density, and wind perturbations. Figure 2c shows such a fishbone structure for the temperature response predicted by theory. The knee of this structure corresponds to the altitude of the center of the body force.
(4) Rayleigh lidar temperature data from the stratosphere and lower mesosphere obtained in June 2014 at McMurdo Station (178°E, 78°S) was analyzed for fishbone structures. Figure 2d shows an example with the knee of such a structure being located at $z \sim 43\,\mathrm{km}$. The corresponding study by Vadas et al. (2018) represents, to the best of our knowledge, the first observational confirmation of the LBF mechanism.
(5) For shorter duration and stronger spatial localization of a body force, the secondary GWs have higher frequencies and larger wavenumbers. In the limiting case of a body force that varies very slowly in time, no secondary GWs are generated. This latter scenario corresponds to the implicit assumptions made in conventional GW schemes (see discussions in Senf and Achatz, 2011; Becker and Vadas, 2020; Bölöni et al., 2021).
(6) The theory of Vadas et al. (2003) and Vadas (2013) describes an idealized picture of the secondary GW generation processes in the real atmosphere that has certain limitations. For example, strong nonlinear interactions of GWs can lead to body-force like perturbations of the ambient flow without the need for a complete dissipation of a primary GW packet (e.g. Fritts et al., 2020; Heale et al., 2022a). Furthermore, when there is a superposition of several primary GW packets (which is most often the case), then the resulting LBFs have smaller spatial and temporal scales than the LBF from a single primary GW packet (Vadas and Crowley, 2010). Hence, the secondary GWs can have smaller scales than the primary GWs. Finally, there is often a superposition of secondary GWs from different LBFs; so it is not always straightforward to identify the secondary GWs in measurements (Vadas et al., 2023a).

3. Gravity-wave spectra and local body-force mechanism

It is commonly believed that GWs having large horizontal wavenumbers and high frequencies are most important for vertical coupling. This notion is rationalized by the following thought experiment. Consider two mid-frequency GWs in the Boussinesq approximation that have the same amplitudes (same energy densities). Then, the GW with the larger vertical group velocity has the larger (absolute) vertical flux of horizontal momentum (hereafter: momentum flux) because 1) the momentum flux is equal to the energy density times k/m, where k and m are the absolute horizontal and vertical wavenumbers, respectively, and 2) the absolute intrinsic frequency is $\omega_I = Nk/m$, where N is the buoyancy frequency. This "equal amplitude" argument suggests that for a given vertical wavelength, GWs having smaller horizontal scales (larger k) or, generally, GWs having larger ω_I should account for larger momentum flux. In the following we show that the observed GW energy spectra suggest that on average, both the GW amplitudes and the GW momentum fluxes are smaller for increasing horizontal wavenumber or for increasing frequency (decreasing horizontal wavelength or decreasing period), thereby negating the relevance of the "equal-amplitude" though experiment.

Observed vertical wavenumber spectra of GWs often show a universal behavior with a m^{-3} exponential slope. This observation can be explained by assuming that all the GWs contributing to the spectrum are at the saturation level (Smith et al., 1987; Fritts and Alexander, 2003). Alternatively, Lindborg (2006) speculated that this universal behavior of observed vertical wavenumber spectra is a reflection of a macro-turbulent inertial range governed by the scaling laws of stratified (macro-)turbulence (hereafter: SMT). Such a cascade would be induced by GW packets that have become dynamically unstable, leading to GW breaking and a cascade to smaller and smaller GWs, which would then also have smaller and smaller periods, until the macro-turbulent cascade transits into Kolmogorov turbulence at the Ozmidov scale (e.g., Avsarkisov et al., 2022). This hypothesis was recently supported by Knobloch et al. (2023) who found the $k^{-5/3}$ spectral behavior in the observed GW horizontal wavenumber spectra. Precisely such a horizontal wavenumber spectrum is predicted by SMT. Our following scaling analysis shows that for SMT, 1) the finding of Smith et al. (1987) is quantitatively consistent with SMT, 2) GW frequency power spectra should have a functional behavior between $\omega^{-5/3}$ and ω^{-2} ($\omega =$ground-based frequency), and 3) the vertical flux of horizontal momentum decreases with increasing horizontal wavenumber and increasing frequency.

SMT assumes a forward energy casacade with regard to the horizontal scales. Hence, the usual scaling analysis from classical turbulence predicts

$$e_k \sim \frac{2}{3}\,\epsilon^{2/3}\,k^{-5/3} \tag{1}$$

for the power spectrum of the horizontal wind, e_k. Here, ϵ is the mechanical dissipation (or frictional heating) rate per unit mass. The scale-dependent aspect ratio for SMT can be written as (Lindborg, 2006; Brune and Becker, 2013)

$$k \;=\; \frac{\epsilon}{N^3}\, m^3 \quad\leftrightarrow\quad m \;=\; N\,\epsilon^{-1/3}\,k^{1/3}\,. \tag{2}$$

Plugging the first equation (2) into (1) and using $e_m = e_k(dk/dm)$, the vertical wavenumber spectrum is

$$e_m \;\sim\; 2\,N^2\,m^{-3}\,, \tag{3}$$

which is often observed at various locations and for ranges of scales (e.g., Chu et al., 2018). Alternatively, to obtain the traditional interpretation of the m^{-3}-spectrum for GWs as proposed by Smith et al. (1987), we assume a spectrum of saturated GWs subject to the mid-frequency and Boussinesq approximations. The saturation condition is $m|T'| = g/c_p$ for the temperature perturbation amplitude T' (Lindzen, 1981). We now use the polarization relations $\omega_I|T'| = (g/c_p)|w'|$ and $k|u'| = m|w'|$, where u' and w' are the GW horizontal and vertical wind perturbations, respectively, as well the GW dispersion relation $\omega_I = Nk/m$. Then the saturation condition can be written as $|u'| \sim N/m$. Furthermore, the integral-scale GW horizontal wind amplitude u_a at wavenumber m is defined as

$$u_a^2 \;\sim\; \int_m^\infty e_{m'}dm'\,. \tag{4}$$

Assuming that u_a fulfills the saturation condition, we get

$$N^2 m^{-2} \;\sim\; \int_m^\infty e_{m'}dm' \quad\rightarrow\quad e_m \;\sim\; 2\,N^2 m^{-3}\,, \tag{5}$$

which is equivalent to Eq. (3). Hence, if an SMT inertial range is governed by the nonlinear interactions of GW modes that fulfill the mid-frequency and Boussinesq approximations, then these GW modes assume integral-scale amplitudes that correspond to the saturation condition.

To obtain the GW frequency spectra in the case of SMT, we assume first that the background wind is zero, which allows us to estimate the intrinsic frequency spectrum. Combining the first Eq. (2) with the dispersion relation, $\omega_I = Nk/m$, yields

$$\omega_I \;=\; \epsilon\,N^{-2}\,m^2 \quad\leftrightarrow\quad m \;=\; N\,\omega_I^{1/2}\epsilon^{-1/2}\,. \tag{6}$$

This expression can be used to transform the vertical wavenumber spectrum (3) into an intrinsic wavenumber spectrum as follows:

$$e_{\omega_I} \;=\; e_m\,(dm/d\omega_I) \;\sim\; \epsilon\;\omega_I^{-2}\,. \tag{7}$$

Since ground-based measurements usually observe GWs subject to Doppler shifting by the background wind such that $\omega_I \geq \omega$ for GWs propagating against the background wind, Eq. (7) is considered to be a lower limit for the observed frequency spectra. The corresponding upper limit is obtained by noting that $\omega_I^3 = (\omega - kU)^3$, where U denotes the (absolute) background wind. Using Eqs. (6) and (2), we get $(\omega - kU)^3 \;=\; \epsilon\,k^2$. Solving for the ground-based frequency yields

$$\omega \;=\; kU + \epsilon^{1/3}k^{2/3} \tag{8}$$

For large-enough k, the first term on the right-hand side of Eq. (8) will become larger than the second term, hence

$$\omega \;=\; kU \quad \text{for finte } U \text{ and large } k. \tag{9}$$

Using Eq. (9), the horizontal wavenumber spectrum (1) can be converted into

$$e_\omega \;=\; e_k\,(dk/d\omega) \;\sim\; \frac{2}{3}\,\epsilon^{2/3}\,U^{2/3}\,\omega^{-5/3}\,. \tag{10}$$

Observed frequency spectra of GWs usually have exponential slopes between $-5/3$ and -2 (e.g., Hoffmann et al., 2010; Guo et al., 2017; Podglajen et al., 2016; Sato et al., 2016; Chen et al., 2016). Equations (7) and (10) prove that these observations are, like the aforementioned result of Knobloch et al. (2023) for GW horizontal wavenumber spectra and the well-known vertical wavenumber spectra, compatible with the hypothesis of a macro-turbulent inertial range. That is, the assumption of SMT combined with the GW dispersion and polarizations relations predicts all these spectra. The assumption of GW saturation, on the other hand, can predict only the vertical wavenumber spectrum.

We now estimate the GW momentum flux spectra that are expected for SMT. Assuming again the mid-frequency and Boussinesq approximation for the GWs, the momentum flux spectra with regard to either k or m are obtained by multiplying the energy spectra (1) or (3) with k/m, where either

m or k is eliminated according to the scale-dependent aspect ratio of SMT (Eq. (2)). The resulting momentum flux spectra are

$$f_k = e_k\, k/m \sim \frac{2}{3}\, \epsilon\, N^{-1}\, k^{-1} \quad \text{and} \quad f_m = e_m\, k/m \sim 2\, \epsilon\, N^{-1}\, m^{-1}\,. \quad (11)$$

These estimates show that the momentum flux decreases for smaller horizontal scales when the scaling laws of SMT apply. Using Eqs. (6) or (9), the momentum flux spectra (11) can be converted into frequency space:

$$f_{\omega_I} = f_m\, (dm/d\omega_I) \sim \epsilon\, N^{-1}\, \omega_I^{-1} \quad \text{and} \quad f_\omega = f_k\, (dk/d\omega) \sim \frac{2}{3}\, \epsilon\, N^{-1}\, \omega^{-1}\,. \quad (12)$$

Hence, the spectral momentum flux decreases with increasing frequency in the case of SMT. Such a behavior has been found in radar observations in the upper mesosphere at high latitudes for periods shorter than a few hours by Sato et al. (2017).

Summarizing, the observed universal behavior of GW spectra for small-enough scales and periods suggests a macro-turbulent inertial range governed by SMT. In particular, the spectral momentum flux decreases with increasing wavenumber and increasing frequency. This analysis supports the relevance of the local body-force (LBF) mechanism described in Sec. 2 since small-scale secondary GWs resulting from the breaking of medium-scale GWs will not contribute efficiently to vertical coupling. Rather, due to the forward cascade associated with the inertial range, these wave modes eventually result in the LBF. The imbalance of the ambient flow then gives rise to vertically propagating secondary GWs that contribute to vertical coupling. Furthermore, medium-scale primary GWs may contribute substantially to the momentum and energy transfer from the troposphere to the middle atmosphere. These GWs are resolved successfully in current high-resolution whole-atmosphere models (e.g., Becker et al., 2022b,a; Vadas et al., 2023a; Liu et al., 2024b).

4. Model description

The Kühlungsborn Mechanistic general Circulation Model (KMCM) was the first GW-resolving general circulation model (GCM) showing that secondary GWs are substantial for understanding the general circulation in the winter mesosphere and lower thermosphere (MLT) (Becker and Vadas, 2018). The HIgh Altitude Mechanistic general Circulation Model (HIAMCM) is a vertical extension of the KMCM with a variety of new components and further developments.

The HIAMCM simulates the neutral dynamics from the surface to the upper thermosphere. It is based on a spectral dynamical core with a terrain-following hybrid vertical coordinate (Simmons and Burridge, 1981). It is currently run with a spectral resolution of T256 (truncation at a total horizontal wavenumber of 256), which corresponds to a horizontal grid-spacing of ~ 52 km and a shortest resolved horizontal wavelength of $\lambda_h \sim 156\,\text{km}$. The effective horizontal resolution is $\lambda_h \sim 200\,\text{km}$ (Becker et al., 2022b). The vertical level spacing is $\sim 600 - 650\,\text{m}$ between the boundary layer and $3 \times 10^{-5}\,\text{hPa}$ ($z \sim 130\,\text{km}$), and increases with altitude above that level, reaching $\sim 10\,\text{km}$ above $z \sim 300\,\text{km}$. Using 280 full layers (L280), the model top is at $4 \times 10^{-9}\,\text{hPa}$, corresponding to $z \sim 450$ km for temperatures of $T \sim 950\,\text{K}$ above $\sim 250\,\text{km}$. The dynamical core is equipped with a correction for non-hydrostatic dynamics and a thermodynamically consistent extension into the thermosphere (Becker and Vadas, 2020).

The HIAMCM includes explicit computations of radiative transfer and water vapor transport, parameterizations of large-scale condensation and moist convection, as well as the full surface energy budget combined with a slab ocean and full topography. Macro-turbulent vertical and horizontal diffusion is parameterized by the classical Smagorinsky model, with the diffusion coefficients depending on the local Richardson number, Ri, giving rise to strong wave damping for $Ri \leq 0.25$ (Becker and Burkhardt, 2007; Becker, 2009). The diffusion scheme accommodates molecular viscosity in the thermosphere for both vertical and horizontal diffusion. As a result, molecular viscosity is the predominant dissipation mechanism for resolved GWs above $z \sim 200\,\text{km}$ (Vadas, 2007; Becker and Vadas, 2020). A simple ion-drag scheme is applied to account for the neutral-ion coupling at low and middle altitudes. The parameterized ion drag furthermore includes a forcing of the auroral circulation in the polar thermosphere (Forbes, 2007; Becker et al., 2022a).

To allow for direct comparison with observational data, the HIAMCM can be nudged to MERRA-2 reanalysis in the troposphere and stratosphere. This nudging is performed in spectral space and is restricted to the planetary-and-synoptic-scale flow. As a result, the explicit simulation of GWs is preserved since GWs are not directly affected by the nudging (Becker et al., 2022b).

The radiation and moist convection schemes are simplified compared to methods used in community models. Furthermore, the HIAMCM does not include a chemistry module, and ion drag is the only ionospheric process that is accounted for. To distinguish these idealizations from methods employed in comprehensive community models, the HIAMCM is called a "mechanistic" model.

5. Evidence of multi-step vertical coupling (MSVC)

5.1. MSVC in the winter stratosphere and MLT

Even though the body-force mechanism for GW generation was proposed in the early 2000s, it took about ten years until the first applications of this mechanism under realistic conditions were published (Vadas and Liu, 2009, 2013; Vadas et al., 2014). These modeling studies focused on secondary GWs in the thermosphere that were generated from the dissipation of convectively generated primary GWs. Later on, the possible fundamental role of MSVC in the winter mesosphere and lower thermosphere (MLT) was suggested by lidar observations of GWs at McMurdo Station in Antarctica (Chu et al., 2011; Chen et al., 2013). These observations showed persistent large-amplitude GWs having medium-to-inertial frequencies and large vertical wavelengths in the winter mesopause region (Chen et al., 2016). In addition, horizontal wavelengths of $\lambda_h \sim 300 - 500\,\text{km}$ in the stratosphere that were inferred by Zhao et al. (2017) were found to be much shorter than the $\lambda_h \sim 800 - 3000\,\text{km}$ estimates by Chen and Chu (2017) for the mesopause region. Such a significant change in λ_h with altitude suggested that the GWs in the mesopause region were secondary, not primary, GWs.

Becker and Vadas (2018) used the free-running KMCM with resolved GWs to simulate the general circulation. They chose several days during June and compared the simulated GWs with lidar observations at McMurdo Station in Antarctica (Chen et al., 2016). Figure 3a shows the observed relative temperature perturbations that were obtained by spectral filtering to retain only periods shorter than 12 hours. The corresponding model result is shown in Fig. 3b. Strong GW amplitudes are seen in the MLT in both plots. These results are unexpected during the wintertime when assuming that primary GWs are the predominant GWs in the middle atmosphere. In particular, only very weak primary GW activity is expected in the winter mesopause region (e.g., Lindzen, 1981; Becker, 2012; Becker and Vadas, 2018). Moreover, the vertical wavelengths of the GWs in Fig. 3 are much longer in the mesopause region than in the stratosphere. This is surprising since during the wintertime, primary GWs propagate predominantly westward in the upper stratosphere and lower mesosphere due to dynamic in-

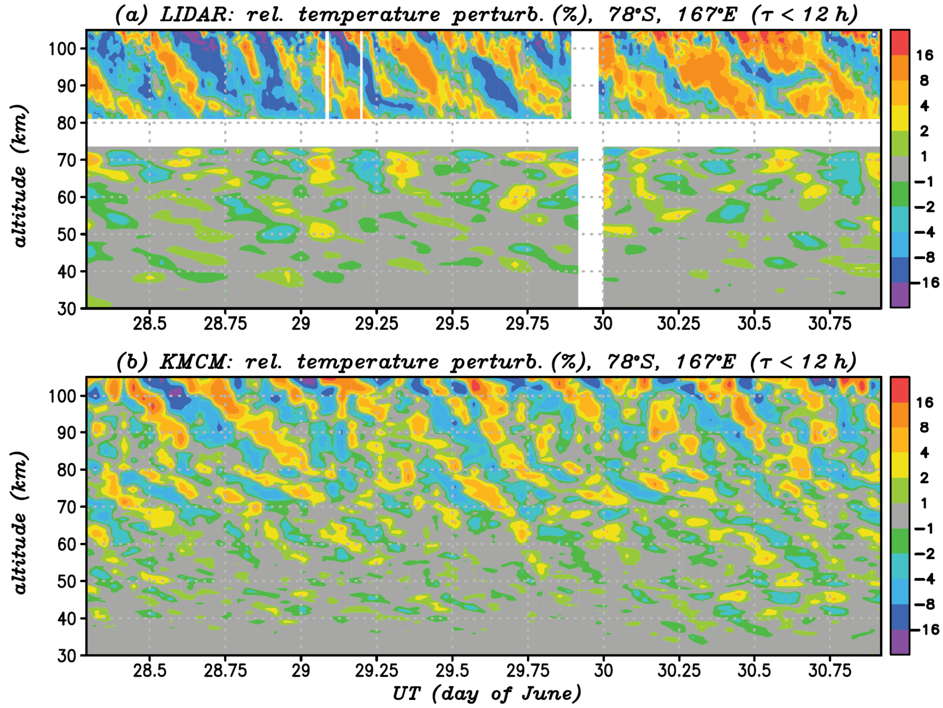

Figure 3: Temperature perturbations at McMurdo (Antarctica) during late June. The perturbations include only periods shorter than 12 hr. (a) From ground-based lidar measurements of Chen et al. (2016). (b) From the KMCM simulation. See Becker and Vadas (2018) for further details.

stability of eastward propagating primary GWs at lower altitudes (Lindzen, 1981; Holton, 1983). Therefore, the vertical wavelengths of the primary GWs would decrease with altitude in the upper mesosphere (rather than increase) since here the strength of the eastward flow associated with the polar vortex decreases with altitude. Hence, the observed and simulated GWs in the mesopause region as shown in Fig. 3 are very likely eastward propagating. Additionally, the simulated horizontal wavelengths in the stratosphere and upper mesosphere were consistent with the aforementioned estimates of Zhao et al. (2017) and Chen and Chu (2017), respectively (Vadas and Becker, 2018). These considerations demonstrate that the GWs observed in the mesopause region are not the same GWs as those that govern the GW field in the stratosphere and lower mesosphere. Rather, most of the GWs at $z > 70\,\mathrm{km}$ during the wintertime at McMurdo Station are most likely secondary GWs.

Further analysis of the model and lidar data showed that the GWs in the southern winter mesopause region were indeed generated in the upper stratosphere and lower mesosphere by the body-force mechanism (Becker and Vadas, 2018; Vadas and Becker, 2018; Vadas et al., 2018). These studies furthermore revealed that the wintertime secondary GWs play a fundamental role for the general circulation. This is illustrated in Fig. 4. The upper row in the figure shows the simulated zonal-mean zonal wind in the southern winter hemisphere from the GW-resolving KMCM and from a corresponding course-resolution model version where orographic and non-orographic GWs were parameterized based on the methods of McFarlane (1987) and Becker and McLandress (2009), respectively (conventional model setup). The overall structure of the zonal wind is quite reasonable and comparable for the two model versions in the stratosphere and lower mesosphere. The GW-resolving simulation exhibits prevailing eastward winds at middle to high latitudes also in the upper mesosphere and lower thermosphere. The conventional model setup, however, produces a wind reversal to westward flow in the upper mesosphere. Such a wind reversal is a general but unrealistic feature of models with parameterized GWs (Smith, 2012). Recently, this deficit of conventional models was further analyzed by Hindley et al. (2022) and Harvey et al. (2022).

The colors in the lower panels in Fig. 4 show the vertical fluxes of zonal momentum per unit mass due to resolved GWs (panel c) and parameterized GWs (panel d). As expected, the westward momentum flux in the conventional model setup is maximum in the stratopause region and decreases to zero with increasing height. When GWs are simulated explicitly, this pattern of the westward momentum flux is reproduced (albeit with somewhat smaller values), but is complemented by a significant eastward momentum flux at higher altitudes. This eastward flux leads to a significant eastward GW drag in the winter polar mesopause region in addition to the usual westward GW drag from primary GWs in the stratosphere and lower mesosphere (contours in Fig. 4c and d). Even though thermal tides and traveling planetary waves also contribute to the wave driving in the mesopause region, it is the eastward GW drag from the secondary GWs that leads to a realistic zonal wind in this regime. Therefore, MSVC is not only important for the interpretation of observed GW perturbations (Fig. 3), but is also an essential new element in our understanding of the general circulation in the MLT.

It may be argued that the eastward GW drag in the winter mesopause region is due to primary GWs having large horizontal phase speeds that were not included in the non-orographic GW scheme of the KMCM (or corre-

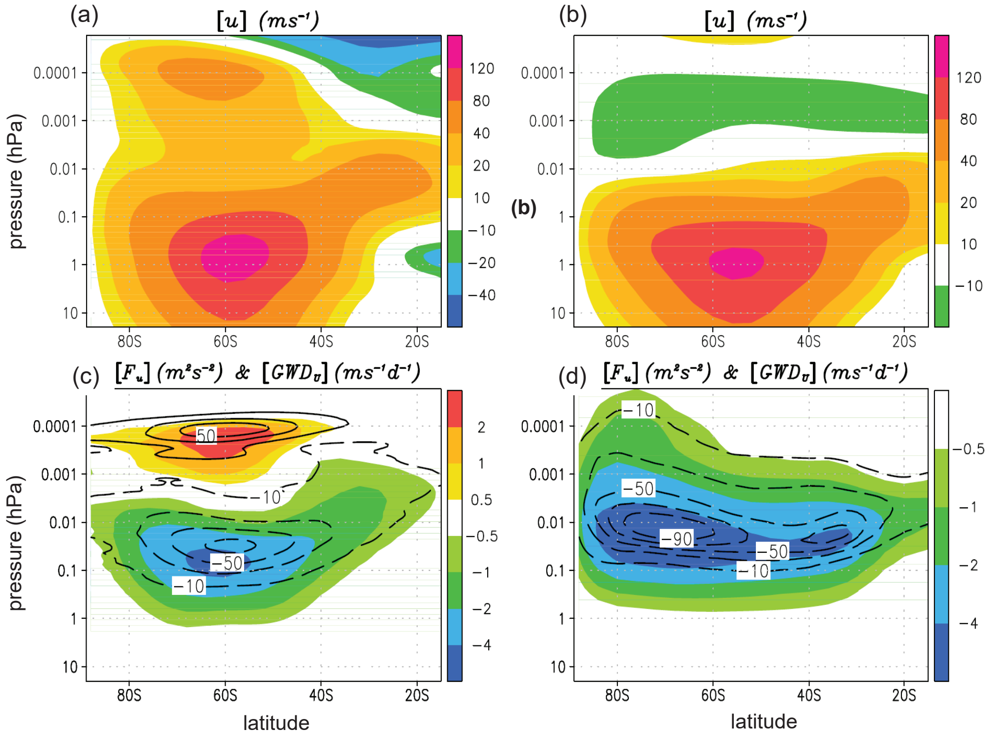

Figure 4: Resolved versus parameterized GWs. Comparison of the high-resolution KMCM with resolved GWs (left column) and a corresponding course-resolution model version with parameterized GWs (right column) in the southern winter hemisphere (averaged from June 21 to August 10). Upper row: Zonal-mean zonal wind. Lower row: Vertical flux of zonal momentum due to GWs (color shading) and GW drag (contours for $0, \pm 30, \pm 50, -70, -90\,\mathrm{m\,s^{-1}d^{-1}}$). For the GW-resolving HIAMCM, GW perturbations are defined by retaining only total horizontal wavenumbers $n > 30$ (horizontal wavelengths $\lambda_h < 1350\,\mathrm{km}$) with respect to the spectral decomposition in terms of the spherical harmonics. The pressure levels correspond to the model's vertical hybrid coordinate times 1013 hPa. See Becker and Vadas (2018) for further details.

sponding GW schemes in other models). However, the southern polar vortex is characterized by large mean winds so that these primary GWs would need to have extremely large eastward phase speeds to avoid critical levels. More importantly, if primary GWs accounted for the eastward GW drag in Fig. 4c, then this effect should be even larger during wintertime in the northern hemisphere since here the polar vortex is much weaker than in the southern winter hemisphere so that more primary GWs having eastward phase speeds and eastward momentum flux would reach the MLT. However, the opposite was simulated with the GW-resolving KMCM. The eastward zonal flow and resolved GW activity in the winter mesopause region are stronger during

July than during January (Becker et al., 2020; Avsarkisov et al., 2022). This hemispheric asymmetry is consistent with satellite observations that also show stronger eastward winds in the winter polar mesopause region during July than during January (Smith, 2012).

Figure 5 shows the zonal-mean circulation for the whole atmosphere from the HIAMCM for 1–20 January 2017 (left column) and 6–20 July 2006 (right column). The HIAMCM simulates reasonably realistic temperatures and zonal winds. This includes the cold summer mesopause and the transition from westward to eastward flow above the temperature minimum, the subtropical mesospheric jet in the winter hemisphere, as well as eastward winds at high latitudes in the winter MLT. The hemispheric differences when comparing July to January include a stronger eastward flow and stronger westward Eliassen-Palm flux (EPF) divergence in the winter stratosphere and lower mesosphere, stronger absolute EPF divergence in the upper mesosphere and a stronger summer-to-winter pole residual circulation and a colder summer polar mesopause. These hemispheric differences are consistent with satellite observations and the interhemispheric coupling mechanism (e.g. Karlsson and Becker, 2016; Körnich and Becker, 2010; Smith, 2012). The eastward flow and eastward EPF divergence in the winter mesopause region is stronger during July than January. According to our discussion above, this hemispheric difference is caused by stronger secondary GW generation when the polar vortex is stronger. For these particular model simulations, also the westward EPF divergence in the summer lower thermosphere is stronger during July. As a result of these hemispheric differences, the reversed residual circulation cell in the lower thermosphere is stronger during July and extends from pole to pole.

The dependence of secondary GWs in the winter upper mesosphere on the polar vortex strength farther below was explained in Vadas and Becker (2019) and Vadas et al. (2024) based on the LBF mechanism. Consider a primary GW packet generated at a height around z_i, propagating upward nearly conservatively, and then dissipating around $z_b > z_i$, thereby creating a LBF at z_b that induces the generation of secondary GWs. The secondary GW amplitudes are proportional to the body force. Considering those secondary GWs at a height $z > z_b$ that propagate nearly conservatively, their amplitudes can be estimated to be proportional to $\exp\{(z_b - z_i)/H\} \times \exp\{(z - z_b/(2H)\} = \exp\{(z + z_b - 2z_i)/(2H)\}$, where H is the density scale height. This equation means that under idealized conditions, the secondary GW generation according to the LBF mechanism works like an amplifier for the GWs at $z > z_b$,

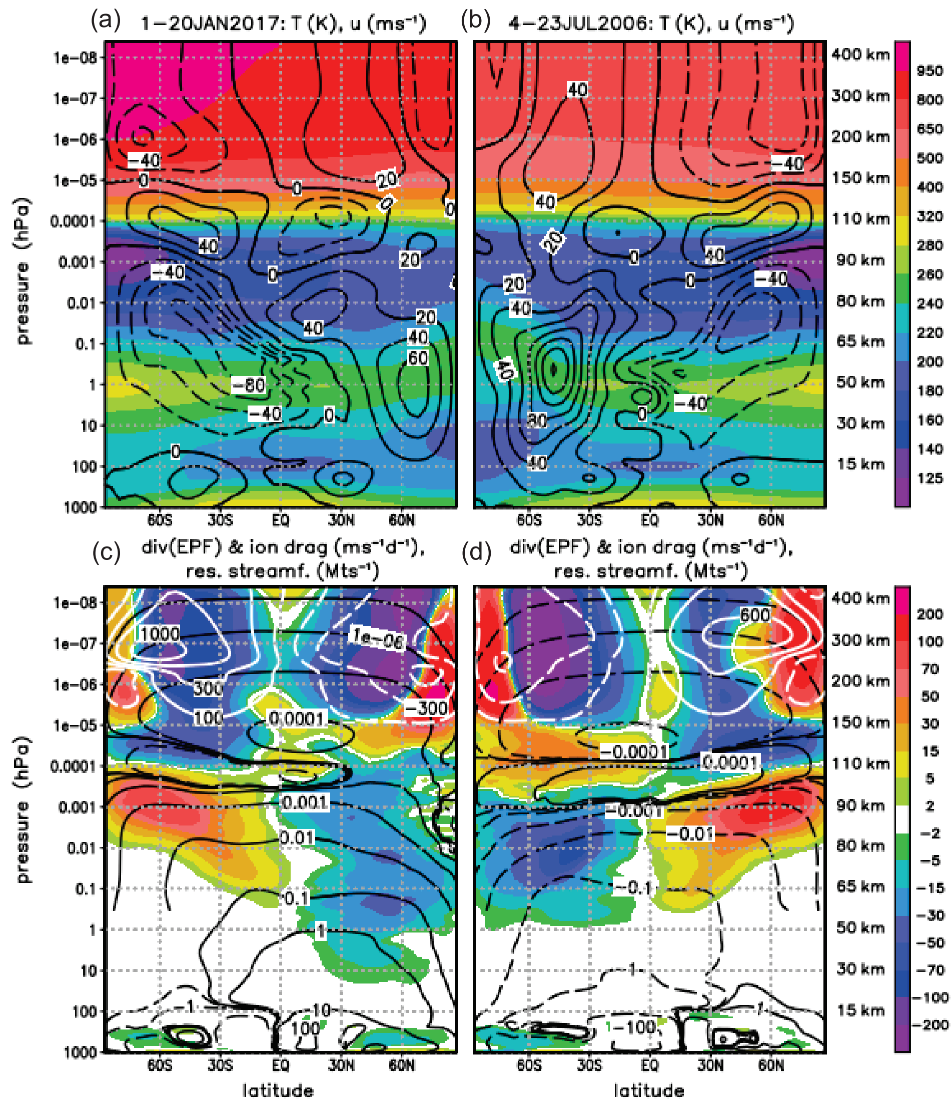

Figure 5: Zonal-mean circulation and wave driving from the HIAMCM for 1-20 January 2017 (left) and 6-20 July 2006 (right). First row: Temperature (colors) and zonal wind (contour interval 20 m/s). Second row: Eliassen-Palm flux (EPF) divergence (colors, unit $\mathrm{m\,s^{-1}d^{-1}}$), residual mass streamfunction (black contours) and zonal component of the parameterized ion drag (white contours). Black contours in (c) are for $10^{-7}, 10^{-6}, \pm 10^{-5}, \pm 10^{-4}, 10^{-3}, 10^{-2}\mathrm{Mt\,s^{-1}}$ above 0.3 hPa and for $\pm 0.1, \pm 1, \pm 10, 100\,\mathrm{Mt\,s^{-1}}$ in the troposphere and stratosphere. Black contours in (d) are as in (c), but with opposite sign. White contours in (c) and (d) are for $\pm 100, \pm 300, \pm 600, 10^3$ $\mathrm{m\,s^{-1}d^{-1}}$. The pressure levels correspond to the model's vertical hybrid coordinate times 1013 hPa. Approximate heights are given on the right-hand sides of panel b and d.

resulting in larger amplitudes when z_b shifts to higher altitudes. Since primary westward GWs break at higher altitudes for a stronger polar vortex, the secondary GWs in winter mesopause region along with their eastward GW drag are also stronger in this case. This simple estimate explains why the EPF flux divergence from about 1×10^{-3} to 3×10^{-5}hPa (about 90 to 120 km) as simulated by the HIAMCM is more eastward in the southern winter hemisphere (Fig. 5d, left half of the panel) than in the northern winter hemisphere (Fig. 5c, right half of the panel). It also explains why the model fidelity in the northern winter MLT of conventional whole-atmosphere models with parameterized GWs (hence, no MSVC) is enhanced during periods of a weak and variable polar vortex (including sudden stratospheric warming events), but is very low when the polar vortex is strong and stable (Harvey et al., 2022). The reason is that secondary (and higher-order) GWs are highly relevant in the latter case and much less so in the former case.

The primary GWs that give rise to MSVC in the winter hemisphere are not necessarily due to only tropospheric sources (e.g., flow over topography and jets and fronts). Recent studies suggest that GWs generated by the polar vortex jet give rise to significant contributions as well (Vadas et al., 2023a, 2024). Figure 6 compares GWs simulated by the HIAMCM with lidar observations at the ALOMAR observatory in northern Norway in January 2016. In this case study, several fishbone structures were observed by the lidar and simulated by the HIAMCM at about the same heights and with about the same timing and amplitudes. As discussed in Sec. 2, such structures are indicative of GW generation from several LBFs at horizontally displaced locations. Analysis of the HIAMCM data showed that the location of the first LBF was about 1500 km farther southeast, and that the primary GWs giving rise to the LBF were generated from the polar vortex jet in the lower stratosphere (Vadas et al., 2023a).

High-resolution direct measurements of the vertical wind, temperature, and metal species in the Antarctic mesopause region in late May 2020 with a Na Doppler lidar profiled the vertical fluxes of sensible heat and meteoric species (Chu et al., 2022). This study found that a significant portion of the observed wintertime GWs propagated downward between $\sim$89 and 95 km. Furthermore, the GW potential energy per unit mass exhibited two local maxima around 85 and 112 km. According to the model study of Vadas and Becker (2019, their Fig. 20), the first and second maximum in the lidar data likely reflected the dissipation of mainly primary and secondary GWs, respectively. Hence, the observed downward propagating GWs between the

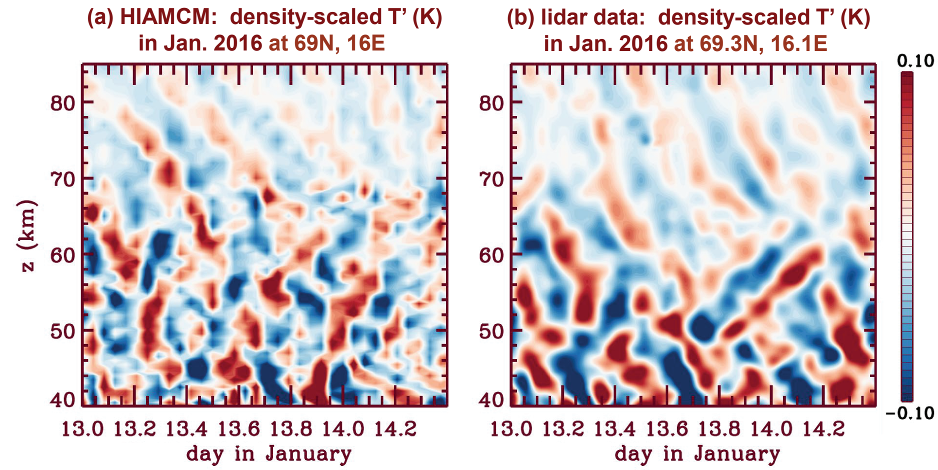

Figure 6: (a) Scaled temperature perturbations $T' \times \exp(-z/2H)$ (H=7 km) on 12-14 January 2016 from the HIAMCM at the site of the ALOMAR observatory in northern Norway. (b) Corresponding result from lidar observations. See Vadas et al. (2023a) for further details.

two energy maxima were likely tertiary GWs generated by the dissipation of secondary GWs at $\sim$112 km. These observations lend further support to the importance of MSVC to explain GW observations in the winter MLT.

Large-amplitude GWs during wintertime in the northern hemisphere at altitudes above $z \sim 85$ km were found in radar observations by Hoffmann et al. (2010, their Fig. 11), but were not identified as secondary GWs at that time. Furthermore, Avsarkisov et al. (2022) computed the annual cycle of the integral turbulent velocity based on radar measurements in the northern hemisphere (see Fig. 10 in their paper). This result reflected the well-known summer-winter asymmetry that is expected for the primary GW amplitudes (e.g., Lindzen, 1981; Becker, 2012), namely a large maximum in the summer mesopause region and a weaker maximum in the winter lower mesosphere. This summer-winter asymmetry is consistent with former rocket-borne measurements of the turbulent dissipation rate by Lübken (1997, see his Figs. 7 and 8). In addition, the result of Avsarkisov et al. (2022) showed a pronounced secondary maximum of the integral turbulent velocity above $z\sim 90$ km during wintertime. This secondary maximum is likely due to secondary GWs. Hence, the results of Hoffmann et al. (2010) and Avsarkisov et al. (2022) confirm that secondary GWs are relevant also in the northern winter hemisphere. It is likely that the rocket-borne measurements of Lübken

(1997) did not reach high enough to capture the dissipation induced by the breaking of the wintertime secondary GWs.

5.2. MSVC in the winter thermosphere and ionosphere

The dependence of MSVC in the winter middle atmosphere on the polar vortex translates into the winter thermosphere and ionosphere. This statement holds particularly true for low geomagnetic activity. A link between the polar vortex and GWs in the thermosphere during geomagnetically quiet times was found by Frissell et al. (2016) and Nayak and Yiğit (2019). These observational studies showed that wintertime traveling ionospheric disturbances (TIDs) at middle latitudes during daytime in the northern hemisphere are correlated with the strength of the polar vortex. Since most quiet-time TIDs are likely caused by GWs from below, this correlation suggests that the strengths of wintertime thermospheric GWs and the polar vortex are positively correlated as well. We therefore expect a pronounced effect when comparing particular periods with a strong polar vortex to periods characterized by a weak polar vortex or even a sudden stratospheric warming (SSW) event. Becker et al. (2022a) addressed the corresponding dynamical mechanism by analyzing a HIAMCM simulation for the winter 2016-2017. This season was characterized by a strong vortex in late December 2016 and an SSW event in late January and early February 2017. In the following we review some results of this study.

Figure 7a,b shows snapshots of GW temperature perturbations and large-scale winds at 60°N and 12 UT on 27 December 2016 (strong vortex period). We use the same wavenumber decomposition based on spherical harmonics to distinguish between the large-scale flow and GWs as in Fig. 4. Panel a indicates strong primary GWs in the stratosphere and lower mesosphere over northern Europe (from about 10°W to 20°E). The phase inclination of these GWs is indicative of westward propagation relative to the mean flow. The flattening of the GW phases in the lower mesosphere over Europe indicates dynamical instability and subsequent dissipation induced by turbulent diffusion, leading to GW-mean flow interaction. This process is induced by the Doppler shifting of the westward GWs towards smaller intrinsic frequencies (shorter vertical wavelengths) caused by westward vertical wind shear (see color in panel b). Similar GW features as over Europe are seen in other longitude bands, for example, from about 90°E to 110°E and from about 140°E to 160°E. In view of the MSVC mechanism discussed above, Fig. 7a suggests that secondary GWs become the predominant GWs in the winter mesosphere for $z \sim 70-100\,\mathrm{km}$. The phase inclinations of these GWs are indicative of both westward and eastward propagation directions, with eastward propagation prevailing with increasing altitude because of westward vertical shear of the large-scale zonal wind (colors in panel b). For example, eastward GWs are visible in the upper mesosphere from about 10°E to 40°E and from about 90°E to 150°E. When the secondary GWs propagate into the mesopause region and lower thermosphere, they become subject to strong refraction by the variable vertical wind shears that are induced by the semi-diurnal tide and traveling planetary waves. Figure 7a indicates that the secondary GWs dissipate in this regime because many of the GW phase lines flatten or become more horizontal with increasing height for $z \sim 90-120\,\mathrm{km}$. Since the resulting wave-mean flow interactions are expected to be localized in space and time, this gives rise to the generation of tertiary GWs. Overall, there are mainly higher-order GWs above about $z \sim 150\,\mathrm{km}$. These waves have very long vertical wavelengths and phase speeds of several 100 $\mathrm{m\,s^{-1}}$ (Vadas, 2007; Vadas and Becker, 2019; Becker and Vadas, 2020), which is much larger than what is typical in the wintertime lower and middle atmosphere. Note that the phase inclinations of the thermospheric GWs in Fig. 7a do not indicate clear westward or eastward propagation directions, especially not in the longitude sector from about 30°W to 60°E.

The lower row in Fig. 7 shows the same snapshots as the upper row, but during the SSW on 31 January 2017. The wind reversal is indicated by the predominantly westward flow in the upper stratosphere and lower mesosphere (colors in panel d). Panel c shows westward and eastward propagating primary GWs in the stratosphere. The phase inclinations of these GWs in the mesosphere indicate predominantly eastward propagation, as is expected due to the predominantly westward large-scale zonal wind. The GWs in the stratosphere and mesosphere during the wind reversal have much weaker amplitudes than during the strong vortex period. The primary GWs dissipate in the upper mesosphere due to eastward vertical wind shear. Again, this process is expected to be localized in space and time, therefore generating secondary GWs that propagate to higher altitudes. The phase inclinations of these secondary GWs indicate eastward and westward propagation directions in the mesopause region and lower thermosphere. Like on December 27, higher-order GWs having large vertical wavelengths emanate from the lower thermosphere. Overall, these GWs have much weaker amplitudes than during the strong vortex period. This is consistent with the aforementioned correlation between TIDs and the polar vortex when assuming that quiet-

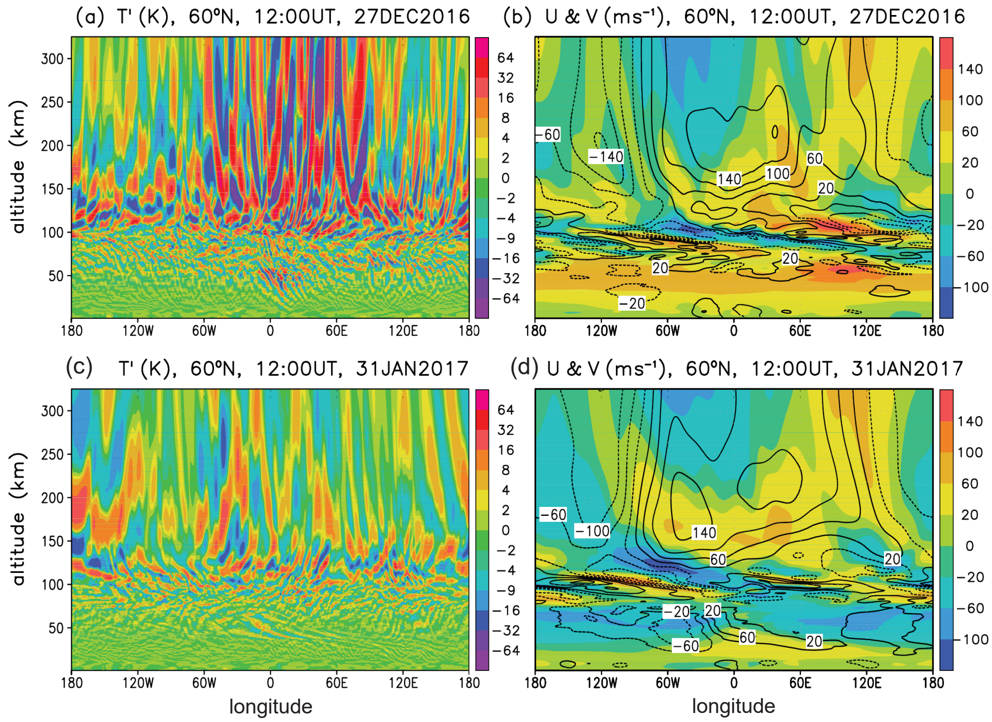

Figure 7: Illustration of MSVC during wintertime as simulated by the HIAMCM. Longitude-height plot of GW temperature perturbations at 60°N on 27 December 2016 (strong vortex) at 12:00 UT. (b) Large-scale zonal wind (colors) and meridional wind (black contours for $\pm 20, \pm 60, \pm 100, \pm 140, 180\,\mathrm{m\,s^{-1}}$. (c) and (d) Same as (a) and (b), but on 31 January 2017 (SSW). See Becker et al. (2022a) for further details.

time TIDs are driven by GWs from below.

The horizontal structures of the higher-order GWs in the thermosphere are illustrated in Fig. 8 by showing north-polar projections of temperature GW perturbations (colors) and the large-scale horizontal wind (white arrows) at $z = 250\,\mathrm{km}$. Assuming that higher-order GWs are generated in the lower thermosphere as concentric ring structures (Vadas and Becker, 2019; Becker et al., 2022a; Vadas et al., 2024), the GW dissipation induced by mainly the diurnal tidal winds leads to partial concentric ring structures at higher altitudes where the largest amplitudes are found for propagation directions that are roughly against the tidal winds. Figure 8a,c indicate three concentric ring structures that are indicative of higher-order GW sources in the lower thermosphere. These sources are located over Scandinavia (visi-

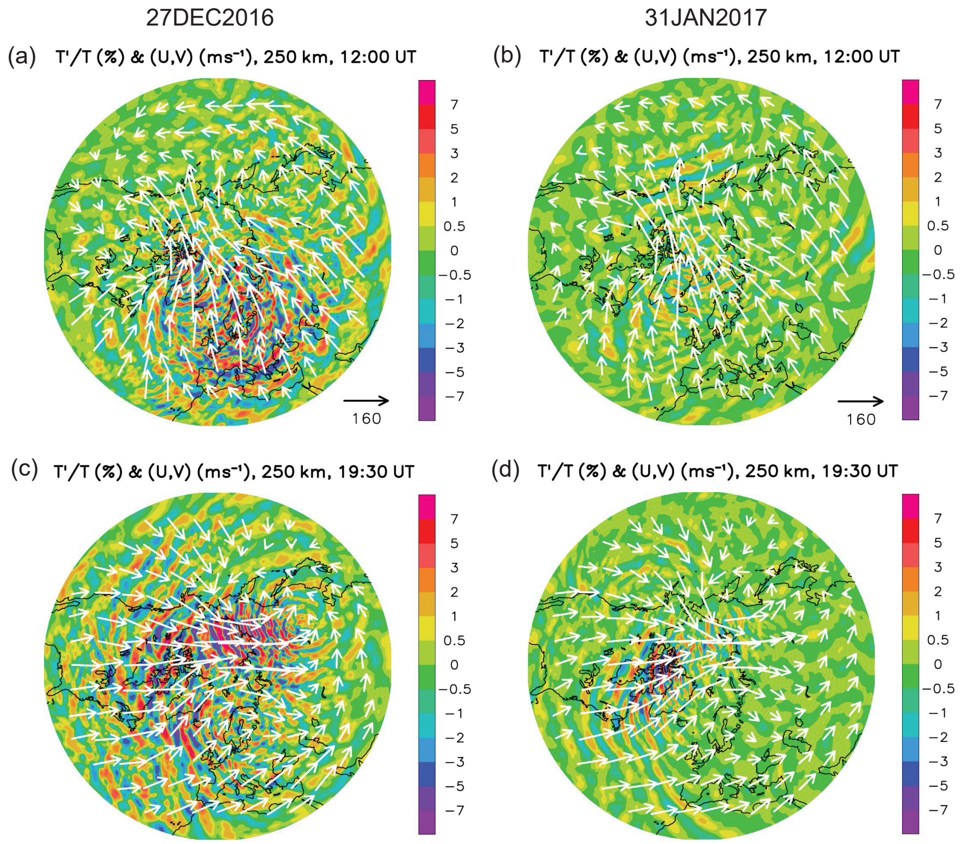

Figure 8: North polar plots of GW temperature perturbations (colors) and the large-scale horizontal wind (white arrows, arrow scale in $\mathrm{m\,s^{-1}}$) at $z = 250\,\mathrm{km}$ from the HIAMCM. (a) 27 December 2016, 12 UT. (b) 31 January 2017, 12:00 UT. (c),(d) Same as (a),(b) but at 19:30 UT. See Becker et al. (2022a) for further details.

ble in panel c), over eastern Siberia (visible in panel a and c), and over the Labrador Sea (visible in panel a). The GWs emanating from these sources propagate mainly equatorward during local noon and early afternoon (over Europe at 12 UT and over North America at 19:30 UT) because the tidal winds are poleward and strongest in these regions. During the SSW event on 31 January 2017, the general behavior of the higher-order GWs is qualitatively similar. Their amplitudes, however, are much weaker as a result of much weaker MSVC as discussed above.

Becker et al. (2022a) determined the average daily cycle of GWs in the thermosphere during the strong vortex period by post-processing the model data from 21 to 30 December 2016. They computed averages from 23:30 to 00:30 UT, 05:30 to 06:30 UT, 11:30 to 12:30 UT, and 17:30 to 18:30 UT, taking all 10 days into account and taking advantage of the 10 minute cadence of the model output. Figure 9 illustrates the resulting average daily cycle in the thermosphere at 10^{-7}hPa ($z \sim 300$ km) at 0, 6, 12, and 18 UT. Colors show the GW kinetic energy per unit mass. The black arrows show the large-scale horizontal wind, while the white arrows show the vertical flux of horizontal momentum due to GWs. The black arrows confirm the daily cycle of the large-scale horizontal wind in the thermosphere that is due to the diurnal tide, with cross-polar flow from the dayside to the nightside, which facilitates equatorward propagation of higher-order GWs during local noon and afternoon (Crowley and Rodrigues, 2012). This average GW propagation direction is evident from the momentum flux vectors (white arrows). These results reveal a significant daily cycle of the GW amplitudes in the winter thermosphere. On average, the GW amplitudes at middle latitudes are strongest where and when the tidal flow is strongest, which is the case roughly around local time noon and afternoon when the tidal flow has a strong poleward component. Furthermore, the GW amplitude maximum extends northwestward, that is, to polar latitudes during local morning. Here, the tidal wind has a strong westward component such that the GWs in this regime exhibit a strong eastward momentum flux. This effect is induced by the auroral circulation (Forbes, 2007), which is further discussed in Becker et al. (2022a).

Figure 10 shows that the simulated GWs in the thermosphere are consistent with observations. The keogram of GW perturbations from the HIAMCM on 30 Decemeber 2016 (panel a) confirms southward propagating GWs at $z = 300$ km. Corresponding perturbations of the total electron content (dTEC) as derived from observations using the Global Navigation Satellite System (GNSS) on the same day (panel b) show a very similar behavior. While there was strong GW and TID activity in late December 2016, these variations had much weaker amplitudes during the SSW in late January 2017, as can be concluded from the simulated GWs and observed dTEC in the lower row of Fig. 10. Moreover, the wave characteristics simulated by the HIAMCM are very similar to those seen in dTEC. This is indicated in the figure by inserting a few phase lines. For case #1 (around local noon on 27 December 2016), these phase lines correspond to southward propagation with a horizontal wavelength of ~ 700 km and a period of ~ 45 minutes, resulting in a ground-based horizontal phase speed of $c_{ph} \sim 270\,\mathrm{m\,s^{-1}}$. The case #2 GW phases occur shortly after nightfall. Becker et al. (2022a) showed that these waves had an approximate southwestward horizontal propagation direction and a period of ~ 60 minutes. The apparent horizontal wavelengths in the

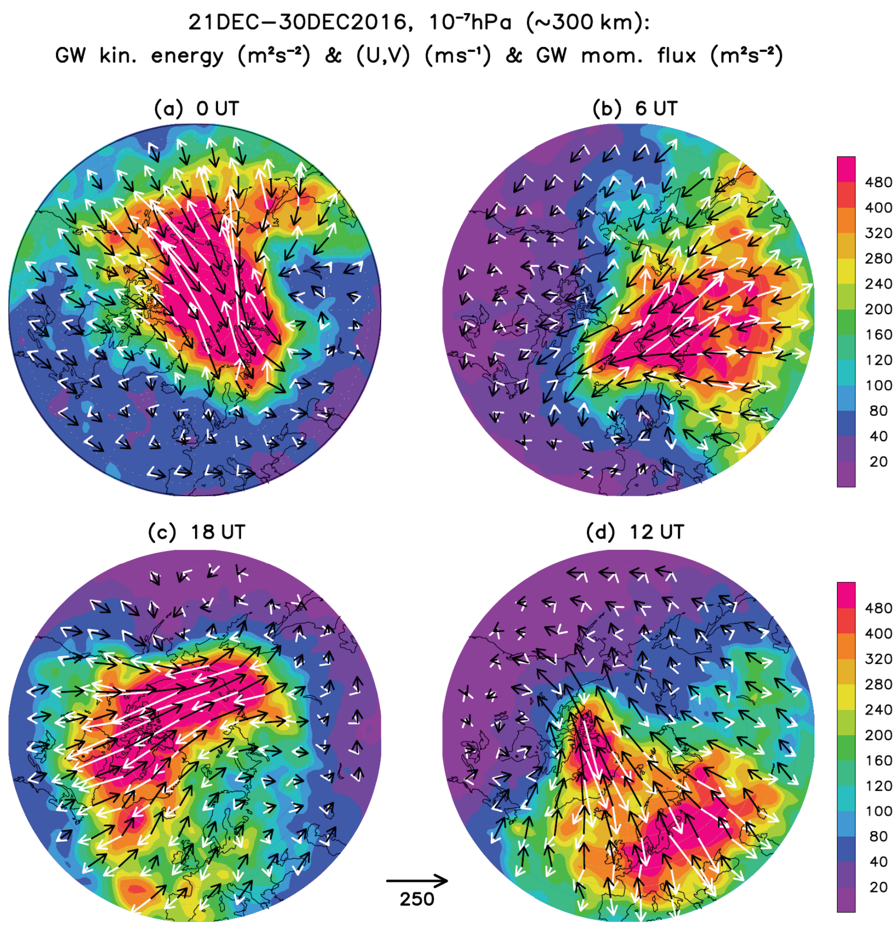

Figure 9: Average daily cycle in the thermosphere during the strong vortex period from 21 to 30 December 2016 at 10^{-7}hPa ($z \sim 300\,\mathrm{km}$). Colors show the average GW kinetic energy per unit mass at (a) 0 UT, (b) 6 UT, (c) 18 UT, and (d) 12 UT. The black and white arrows show the corresponding average large-scale horizontal wind and the vertical GW flux of horizontal momentum, respectively. The arrow scale (between panel c and d) is the same for the black and white arrows, but refers to the respective units, that is $\mathrm{m\,s^{-1}}$ or $\mathrm{m^2s^{-1}}$. See Becker et al. (2022a) for further details.

longitudinal and latitudinal directions were estimated to be 800 and 900 km, respectively, yielding a true horizontal wavelength of $\sim 450\,\mathrm{km}$ and a phase speed of $c_{ph} \sim 130\,\mathrm{m\,s^{-1}}$. Such values are typically observed for medium-to-large-scale TIDs, even though the simulated wavelengths are somewhat

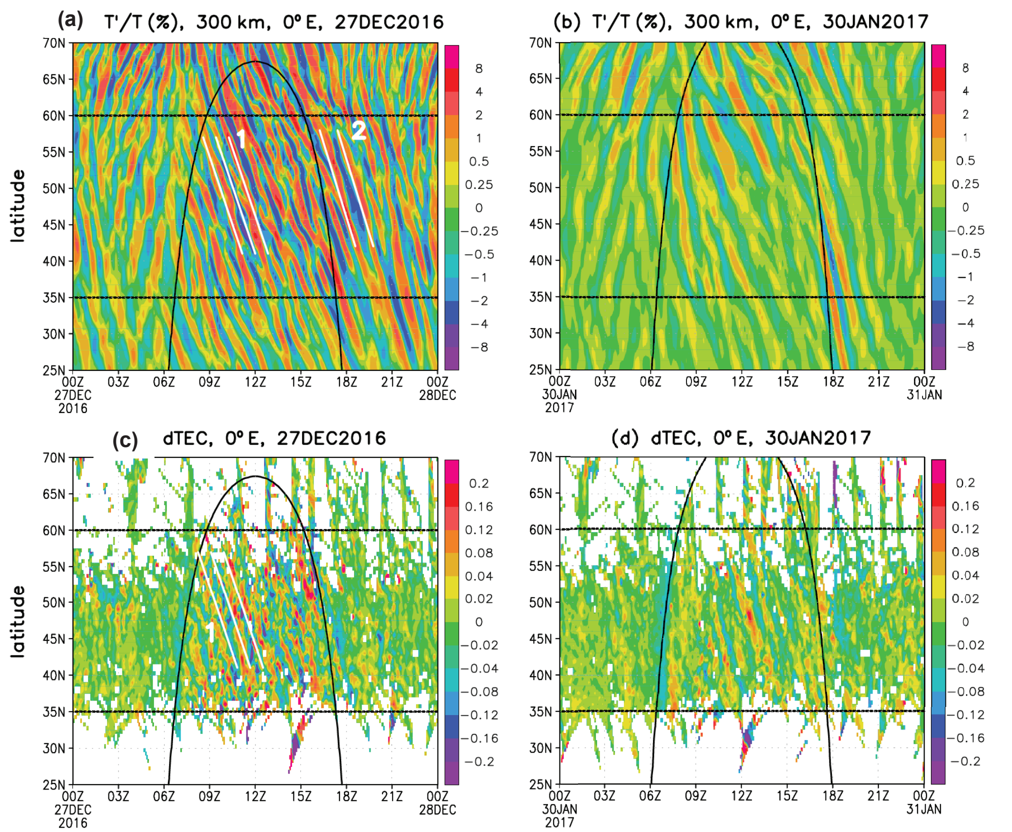

Figure 10: Keograms of the simulated relative temperature perturbations at 300 km height (upper row) and observed dTEC (lower row) at 0°E and from 25°N to 70°N. The left column is for 27 December 2016 (strong vortex) and the right columns is for 30 January 30 (SSW). The straight white lines are inserted to indicate wave phases (see text). Sunrise and sunset are indicated by the curved black lines. See Becker et al. (2022a) for further details.

larger than found by Frissell et al. (2016). Figure 10 strongly suggests that the observed quiet-time TIDs are driven to a large extent by GWs that result from MSVC. Becker et al. (2022a) mentioned that the observed TID activity amplified again from about 31 January to 7 February 2017 despite the ongoing SSW. This amplification was very likely due to geomagnetic forcing (Kp=3+ to Kp=5 during this period). Such an effect was not reflected by the GWs simulated by the HIAMCM due to the lack of geomagnetic forcing in the model.

Figure 9 implies that there is a strong cancellation of the GW momentum flux and drag contributions from different longitudes in the zonal and temporal mean because of the variations of GW propagation and dissipation induced by the tidal winds. Therefore, a zonal-mean zonal GW drag obscures most of the information about the actual GW-mean flow interactions. Indeed, GW-mean flow interaction in the winter thermosphere is mainly due to GW-tidal interaction. The crucial role of the GW-tidal interactions is also well known for the MLT (e.g. Senf and Achatz, 2011; Becker, 2017; Becker and Vadas, 2018; Heale et al., 2022a; Becker and Oberheide, 2023). Many details of these GW-tidal interactions remain to be understood.

The HIAMCM results presented in Becker et al. (2022a) indicate that the GW drag in the zonal direction is not very relevant in the thermosphere when compared to the contributions from thermal tides and ion drag. Even though community whole-atmosphere models with parameterized GWs do not include MSVC, they also do not predict any notable effect in the thermosphere from parameterized GWs (e.g. Fomichev et al., 2002; Liu et al., 2009; Smith, 2012). The reason is that primary GWs from the mid-latitude troposphere that propagate to the thermosphere are efficiently dissipated by molecular viscosity (Vadas, 2007; Becker and McLandress, 2009). Indeed, the strong EPF divergences simulated in the thermosphere above about 150 km height (see Fig. 5) are mainly due to tides. Nevertheless, the contributions from mainly secondary GWs in the lower thermosphere are still relevant in the zonal mean and contribute to the driving of the reversed residual circulation cells, including their hemispheric difference as discussed in Becker and Oberheide (2023). The diurnal thermal tide in the thermosphere is mainly forced (generated) by solar heating in the $z \sim 150 - 200\,\mathrm{km}$ height regime (Torr et al., 1981), but in addition by ion drag forcing in the polar regions (Forbes, 2007). The latter effect leads to the eastward EPF divergence in the polar thermosphere in Fig. 5c,d as discussed in Becker et al. (2022a). In addition, the zonal component of the ion drag (white contours in Fig. 5a,c is the major driver of the summer-to-winter circulation in the upper thermosphere as was recently confirmed by Liu et al. (2024a).

5.3. Secondary GWs from the Hunga Tonga-Hunga Ha'apai volcanic eruption

Even the highest resolutions currently feasible in GW-resolving whole-atmosphere models are not sufficient to simulate the primary GWs from deep convection or volcanic eruptions. If MSVC from such primary GWs needs to be taken into account, the localized and intermittent ambient-flow effects that generate secondary GWs at higher altitudes need to be precalculated by other means and then implemented into the global model. Vadas et al. (2023b,c) and Huba et al. (2023) performed these precalculations of small-scale, high-frequency GWs using a combination of tools that is called the Model for gravity wavE SOurce, Ray trAcing and reConstruction (MESORAC).

The MESORAC was applied to simulate the thermospheric and ionospheric disturbances caused by the eruption of the Hunga Tonga-Hunga Ha'apai (hereafter: Tonga) volcanic eruption on 15 January 2022. To this end, the primary GWs from the Tonga event were inferred from the updrafts in the stratosphere and mesosphere as observed by NOAA's Geostationary Operational Environmental Satellite (GOES). These GWs were ray-traced forward in time, and the GW field was reconstructed using the GW phases and the GW dissipative dispersion and polarization relations. The background atmosphere for this ray-tracing computation was taken from the HIAMCM simulation for 15 January 2022 without any perturbations related to the Tonga event ("no-Tonga" run). The ray-traced primary GWs dissipated from both saturation and molecular viscosity. This resulted in very strong LBFs (and heatings). We obtained localized and intermittent accelerations of more than $1\,\mathrm{m\,s^{-2}}$. The majority of these effects occurred about 30-90 minutes after the major eruption and were confined to the geographical region around the volcano. That is, the primary GWs from the Tonga event could not account directly for any of the far-field effects that were observed around the globe after the volcanic eruption. The HIAMCM simulation of the January 15-19 period was then repeated with the precalculated LBFs and heatings from MESORAC ("Tonga" run). The HIAMCM responded with a broad spectrum of large-amplitude, high-phase-speed secondary GWs that propagated around the globe for several days in the thermosphere. Figure 11a shows an example of a LBF from the dissipation of primary GWs computed with the MESORAC. Note that this LBF occurred at about 5:30 UT in the lower thermosphere only a few hundred kilometers away from the volcano and about one hour after the first eruption. Such LBFs generated secondary GWs in the HIAMCM. Figure 11b shows a snapshot of these secondary GWs in terms of the vertical wind at 280 km and 12:00 UT on 15 January 2022.

These secondary GWs can be compared to satellite observations of winds by ICON-MIGHTI (Harding et al., 2017; Immel et al., 2018). Figure 12 shows the ICON-MIGHTI zonal winds over the eastern Pacific, North America, and the Atlantic Ocean. These winds display the GWs from Tonga

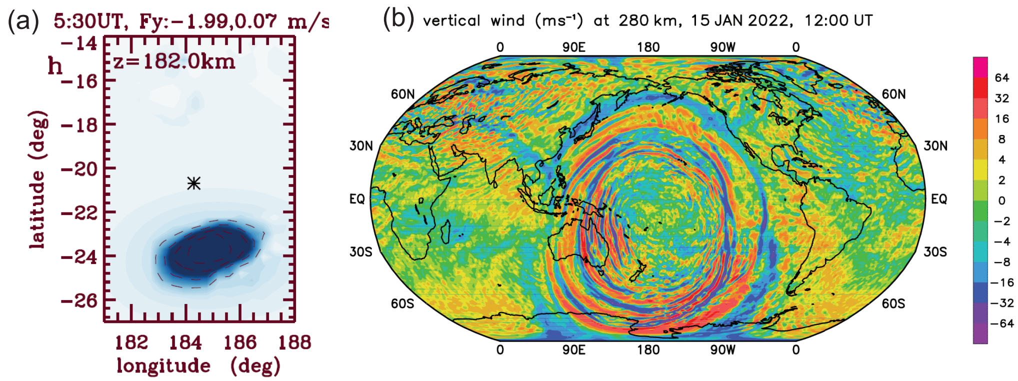

Figure 11: (a) Example of a southward local body force (LBF) at $z = 182$ km and 05:30 UT on January 15, 2022. Such LBFs were computed with the MESORAC from the dissipation of primary GWs that were generated by the updrafts from the Tonga volcanic eruption (which primarily occurred from about 04:15 to 05:50 UT). (b) Snapshot of the secondary GWs simulated by the HIAMCM in terms of the vertical wind at $z = 280$ km and 12:00 UT. See Vadas et al. (2023b) for further details.

during four consecutive orbits. The left column shows the tangent longitude versus latitude and time during these orbits. The right column shows the measured zonal winds. We included pink vertical lines that mark the locations of disturbances that would have propagated from the region of the Tonga eruption with various horizonal phase speeds. Overall these plots indicate that the GWs from Tonga had very large amplitudes, large scales, and horizontal phase speeds of 100 to 600 $\mathrm{m\,s^{-1}}$. Figure 13 compares the analyzed ICON-MIGHTI zonal winds with HIAMCM zonal winds sampled along the same orbits. Direct comparisons showed that the timing and the amplitudes of the HIAMCM zonal wind perturbations agree reasonably well with ICON-MIGHTI (Vadas et al., 2023b). To achieve the best agreement, however, we extracted the zonal wind perturbation amplitudes related to the Tonga eruption by computing the difference between the Tonga run and the no-Tonga run, and then multiplied these wind perturbations by a factor of 1.5. The "scaled Tonga wind" is given by the no-Tonga winds plus the scaled perturbations. In the right column of Fig. 13 we plot these model data 30 minutes later than the satellite data. Eastward and upward-propagating GWs are observed west of 0°E on the dayside. The along-track wavelengths increase with distance from Tonga, as is expected for GWs generated by a point source. In addition, the GWs in each successive orbit have smaller amplitudes and smaller phase speeds, which is expected as well.

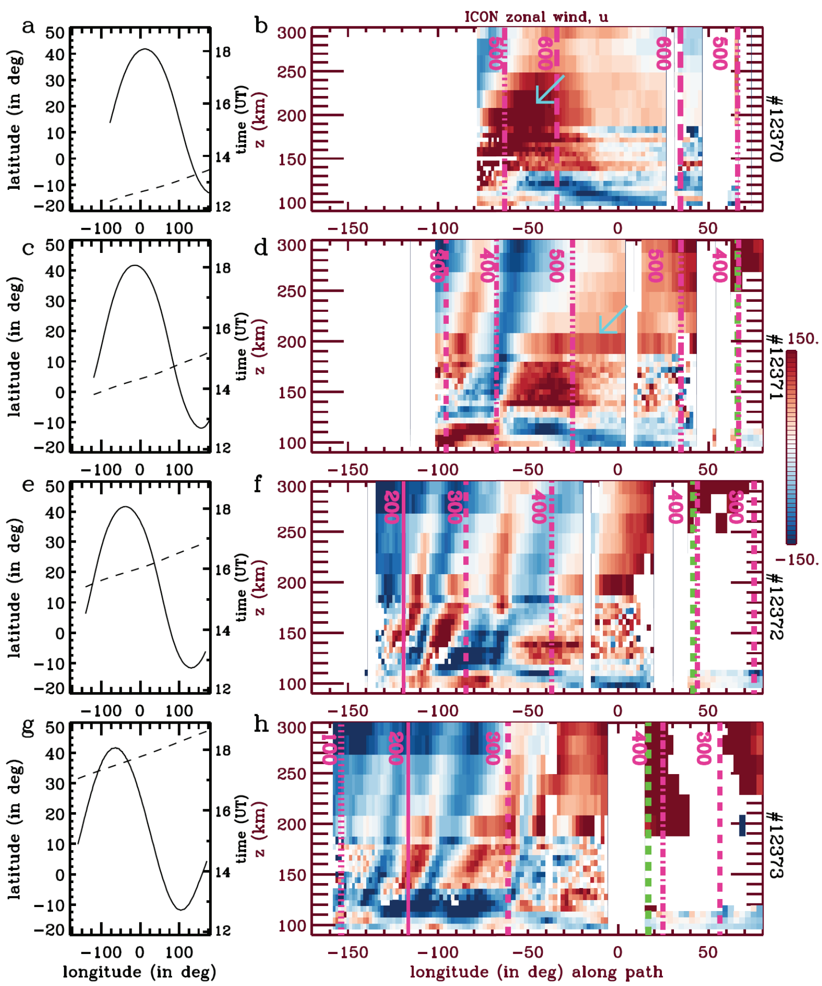

Figure 12: (a) ICON-MIGHTI tangent longitude versus latitude (solid line, left y-axis) and tangent longitude versus time (dashed line, right y-axis) for orbit #12370 on 15 January 2022. (b) ICON-MIGHTI zonal wind as a function of the tangent longitude for orbit #12370. Rows 2-4: Same as row 1 but for orbits #12371, 12372 and 12373. The vertical pink lines show the locations of waves that originated in the thermosphere above Tonga at 5:00 UT with horizontal phase speeds of $100, 200, 300, 400, 500$, and $600\,\mathrm{m\,s^{-1}}$, as labelled. Phase-speed lines are not shown where the westward and eastward waves from Tonga would overlap. Turquoise arrows indicate the fastest large-scale secondary GWs with horizontal phase speeds larger than $500\,\mathrm{m\,s^{-1}}$. Green-dashed lines show the solar (sunset) terminator. From Vadas et al. (2023b).

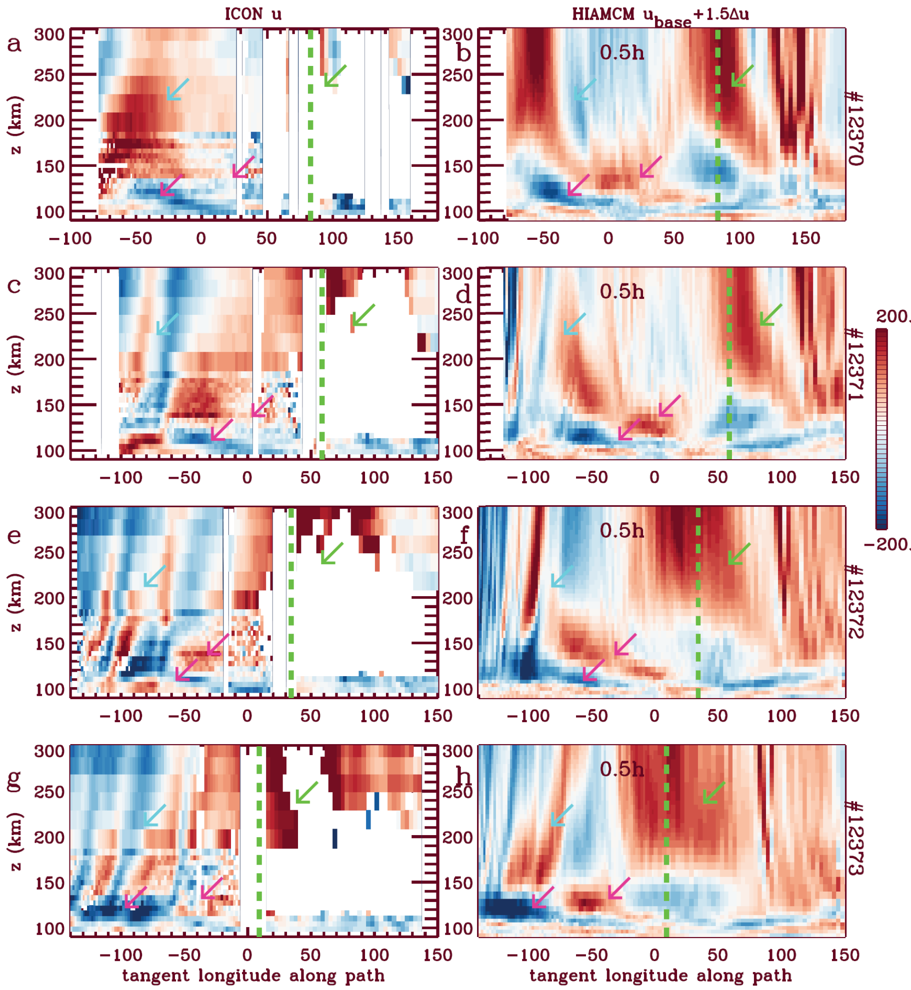

Figure 13: (a),(c),(e),(g) Zonal winds measured by ICON as functions of the tangent longitude and height z for orbits #12370, 12371, 12372, and 12373, respectively. (b),(d),(f),(h) HIAMCM zonal winds from the no-Tonga run plus 1.5 times the perturbation from the Tonga minus the no-Tonga run as functions of the tangent longitude and z and sampled 30 minutes later than the ICON times for orbits #12370, 12371, 12372 and 12373, respectively. The HIAMCM winds are smoothed over 200 km horizontally prior to sampling. The green dashed lines show the sunset solar terminator. Turquoise, pink and green arrows indicate the Tonga GWs, the tides, and the terminator waves, respectively. These arrows are in the same locations in each row. From Vadas et al. (2023b, their Fig. 5).

Summarizing, the ICON-MIGHTI instrument observed northeastward-propagating secondary GWs from Tonga having horizontal phase speeds of about 100 to 600 $\mathrm{m\,s^{-2}}$ and horizontal wavelengths of about 800 to 7500 km, which is in good agreement with the model results. Also the timing and the very large amplitudes of the observed waves are in reasonable agreement with the model. Thus, these observations provide excellent confirmation that the LBF mechanism applies as an explanation for the (secondary) GWs from the Tonga eruption.

To further validate the HIAMCM simulation of the secondary thermospheric GWs from the Tonga event, the model output was used to drive the SAMI3 ionospheric model (Huba et al., 2023). The keograms of the dTEC as simulated by the SAMI3 (Vadas et al., 2023c) were similar to those published by Themens et al. (2022). These authors analyzed the dTEC for a number of regions and computed keograms of the dTEC along great circles from Tonga through these regions (Themens et al., 2022, their Fig. 4). They used a detrend window of 30 minutes for this analysis. Vadas et al. (2023c) revisited these data and used a detrend window of 1 hr to better emphasize the larger-scale secondary GWs (see Fig. 7 in that study). The model output from the SAMI3 was then analyzed in the same way and the result is shown here in Fig. 14. Vadas et al. (2023c) found that the SAMI3 reproduced the GNSS dTEC observations very well. This holds particularly for the wave periods and for the different arrival times of the wave perturbations in different regions. Figure 14 also shows that the secondary GWs with longer periods arrive first, implying that these waves have larger horizontal phase speeds and larger vertical group velocities. This feature of the secondary GWs is typical for a GW spectrum initiated by the LBF mechanism (Vadas et al., 2018; Vadas and Becker, 2019), which is another confirmation that this mechanism gives a consistent picture for the wave perturbations in the thermosphere that were caused by the Tonga volcanic eruption. Note that the fastest (longest-period) secondary FWs were also observed to have decreased the F region peak by $\sim 120\,\mathrm{km}$ over the western US (Vadas et al., 2023c).

Alternative attempts to explain the wave perturbations in the thermosphere from the Tonga event proposed that the observations were due to Lamb waves or Lamb waves leaking into the thermosphere as GWs (e.g. Zhang et al., 2022; Liu et al., 2023). Vadas et al. (2023c) showed that such an interpretation is inconsistent with the observations, especially with those very-large-amplitude GWs that were observed worldwide and had horizontal phase speeds $> 300\,\mathrm{m\,s^{-1}}$. On the other hand, MSVC as simulated by our combination of models explains most observations reasonably well and in a dynamically consistent fashion.

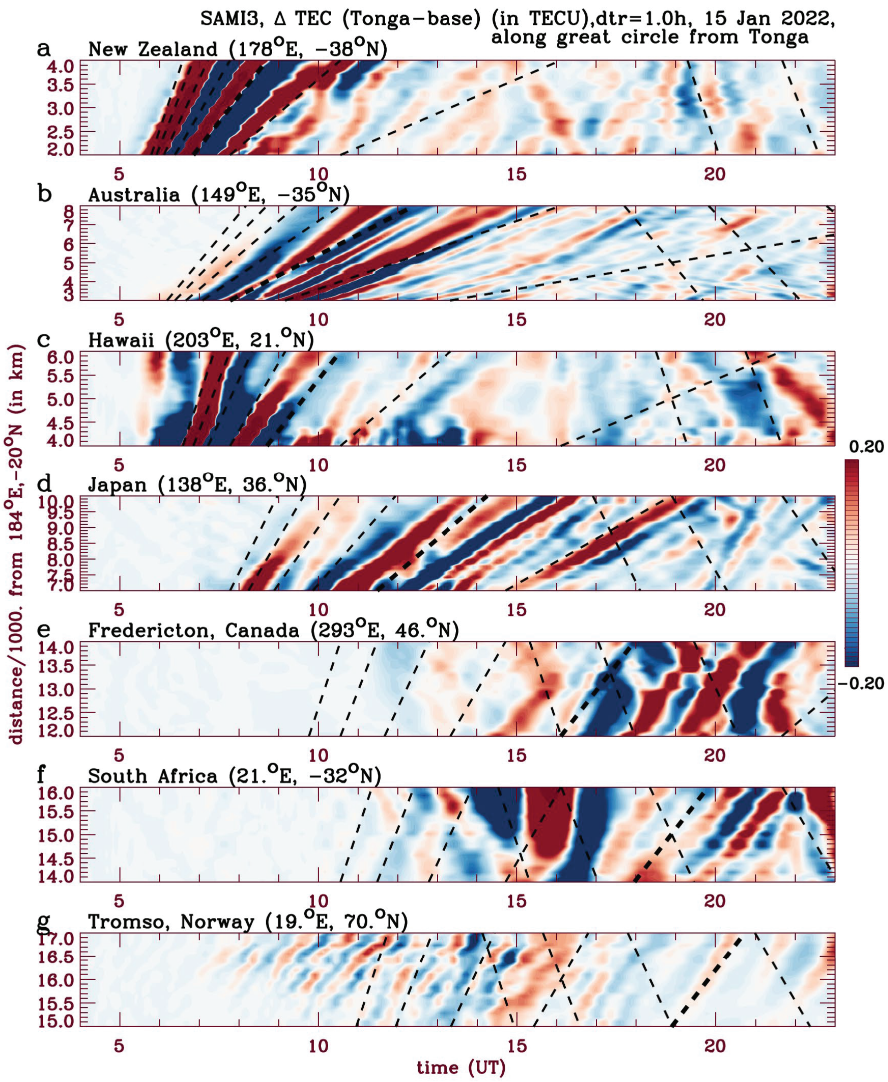

Figure 14: Keograms of the SAMI3 dTEC (in TECU) on 15 January 2022 for the Tonga run minus the base run along great circles from Tonga to (a) New Zealand (178°E, 38°S); (b) Australia (149°E, 35°S); (c) Hawaii (157°W, 21°N); (d) Japan (138°E, 36°N); (e) Fredericton, Canada (67°W, 46°N); (f) South Africa (21°E, 32°S); (g) Tromsø, Norway (19°E, 70°N). A detrend window of 1 hr is applied. Dashed lines show the horizontal phase speeds of outbound and inbound thermospheric GWs that originated above Tonga at 5:00 UT in decreasing steps of 100 $\mathrm{m\,s^{-1}}$, beginning on the left (for the outbound GWs) at 700 $\mathrm{m\,s^{-1}}$. The 500 $\mathrm{m\,s^{-1}}$ dashed phase speed line is darker. The colors are oversaturated to emphasize the waves. From Vadas et al. (2023c).

6. Summary and some open questions

We reviewed the multi-step vertical coupling (MSVC) due to GWs and the relevance of this mechanism to the general circulation and to GWs in the upper mesosphere and thermosphere/ionopshere. At the heart of MSVC is the local body-force (LBF) mechanism according to which the localized (in space and time) breaking or dissipation of GW packets leads to imbalances in the ambient flow which in turn result in the generation of new (secondary or higher-order) GWs (Vadas and Fritts, 2002; Vadas et al., 2003, 2018). We distinguished this GW generation from the energy cascade associated with GW breaking that is characterized by energy and momentum flux transfer to smaller and smaller scales. We argued that small-scale, high-frequency GWs observed in the middle atmosphere that are subject to the universal behavior of GW spectra (energy spectra with spectral slopes of m^{-3} and ω^{-2} to $\omega^{-5/3}$ with regard to the vertical wavenumber and frequency, respectively), are likely a reflection of stratified macro-turbulence (SMT) that is maintained by the instability and breaking of incident GW packets. We showed that observed vertical wavenumber, horizontal wavenumber, and frequency spectra of GWs are all consistent with such a macro-turbulent inertial range on average for small-enough scales and periods. Note, however, that high-frequency primary GWs from deep moist convection or volcanic eruptions do not fit into this category. Such primary GWs often have long vertical wavelengths and correspondingly large vertical group velocities so that they can propagate to the mesosphere or thermosphere before becoming unstable or being dissipated directly by molecular viscosity (Vadas, 2007; Vadas and Liu, 2013; Vadas et al., 2014). It is an ongoing research question as to whether primary or secondary GWs from deep moist convection play the major role for GWs/TIDs in the thermosphere/ionosphere above regions of deep moist convection (see Vadas and Azeem, 2021; Heale et al., 2022b).

We presented a number of examples of MSVC that were simulated with the High Altitude Mechanistic general Ciculation Model (HIAMCM) and a former model version, and which were compared to observations for validation. According to MSVC, the observed large-amplitude GWs in the winter mesopause region at middle to high latitudes (e.g., Chen et al., 2016) are secondary GWs that are generated around and above the wind maximum of the polar night jet. Importantly, these secondary GWs are eastward propagating in the winter mesopause region (Becker and Vadas, 2018) and are essential for understanding the observed prevailing winds in the winter mesosphere and lower thermosphere (Hindley et al., 2022). To date, only GW-resolving models can simulate these GWs and their eastward GW drag in the winter mesopause region. Furthermore, the eastward secondary GW drag is stronger for a stronger polar vortex (Vadas and Becker, 2019; Becker et al., 2022a; Harvey et al., 2022).

The dissipation of the wintertime secondary GWs in the strong wind shears associated with thermal tides and traveling planetary waves in the lower thermosphere leads to the generation of tertiary (higher-order) GWs, which again can be understood by the LBF mechanism (Vadas and Becker, 2019; Vadas et al., 2023a, 2024). Therefore, GWs in the winter thermosphere are mainly higher-order GWs. Model results show that these GWs undergo a strong daily cycle, with GW amplitudes being maximum on average around local noon and early afternoon at F region altitudes (Becker et al., 2022a). These wintertime GWs propagate equatorward (against the tidal flow), confirming that most of the quiet-time traveling ionospheric disturbances (TIDs) during the daytime are driven by GWs from below. In contrast to conventional wisdom, however, the GWs that drive wintertime TIDs are by no means primary GWs from the troposphere, but are higher-order GWs resulting from MSVC. Moreover, since the generation of secondary and higher-order GWs is correlated with the strength of the polar vortex (Vadas and Becker, 2019), MSVC also explains why the strength of observed quiet-time TIDs is correlated with the strength of the polar vortex (Frissell et al., 2016; Nayak and Yiğit, 2019; Becker et al., 2022a).

Global-modeling studies of MSVC from deep moist convection were performed by Vadas and Liu (2009, 2013) and Vadas et al. (2014). These studies used a global model with an effective horizontal resolution $\lambda_h \sim 2000$–3000 km to simulate large-scale secondary GWs from deep convection. However, medium-scale secondary GWs should also have been generated, as was shown by reverse ray-tracing (Vadas and Crowley, 2010), which then would have induced medium-scale TIDs. These earlier model studies remain to be revisited using a global model with high-enough effective resolution to capture all the relevant secondary GWs. The HIAMCM or the latest version of the Whole Atmosphere Community Climate Model with ionosphere eXtension (WACCM-X) (Liu et al., 2024b) would be such global model candidates. In principle, the generation of primary GWs by deep moist processes is explicitly included in such models (e.g. Liu et al., 2014). However, since these models cannot adequately resolve this process due to insufficient resolution, the explicitly simulated MSVC from parameterized deep moist convection is not expected to be realistic as compared to what has been achieved for wintertime conditions for the primary GWs from jets and fronts, flow over topography, and the polar vortex (e.g. Becker et al., 2022b; Vadas et al., 2023a, 2024).

To adequately include the primary and secondary GWs from deep moist convection in a GW-resolving whole-atmosphere models, such models need to be coupled to local convection models like the Model for gravity wavE SOurce, Ray trAcing and reConstruction (MESORAC). The GWs from both models then need to be coupled to an ionospheric model (SAMI3) to elucidate the effects of MSVC from deep moist convection on the thermosphere/ionosphere. In this paper we reviewed simulation results where the MESORAC model suite was applied to compute the primary GWs and LBFs in the middle and upper atmosphere that were generated by the Tonga volcanic eruption on 15 January 2022. The LBFs were used to perturb the HIAMCM to simulate the secondary GWs induced by the Tonga event (Vadas et al., 2023b). The HIAMCM output was furthermore used to drive the ionospheric model SAMI3 to simulate the corresponding ionospheric disturbances (e.g. Huba et al., 2023). These model results showed very good agreement with satellite and GNSS observations of thermospheric waves and ionospheric disturbances caused by the Tonga event, thereby validating the coupling of the three models. A full coupling of the MESORAC to a whole-atmosphere model to simulate MSVC from deep moist convection routinely is yet to be developed. Such a coupled model would be particularly useful to investigate the role of MSVC in the tropics, as well as from subtropical to middle latitudes during summertime.

The simulation of MSVC in high-resolution models needs to be further improved in several ways. For example, Chu et al. (2022) analyzed lidar data in the southern winter MLT and discovered upward GW heat (temperature) flux in a layer around $\sim$97-106 km. Such a flux cannot exist when GWs are considered in the anelastic approximation, where the heat flux is either zero in the conservative case or downward in the case of thermal dissipation (e.g., Becker, 2017). On the other hand, when taking the fully compressible polarization relations of Vadas (2013, Appendix B) into account, upward (downward) conservatively propagating GWs possess an upward (downward) heat flux. The upward heat flux observed by Chu et al. (2022) was presumably due to upward propagating secondary GWs. This newly discovered phenomenon of MSVC remains to be simulated with GW-resolving circulation models. Regarding current community whole-atmosphere models with parameterized

GWs we note that these models cannot simulate MSVC at all. The reason is that all routinely used conventional GW schemes are based on the single-column and steady-state approximation which exclude MSVC by definition (see discussions in Becker and Vadas, 2020; Achatz et al., 2024). New GW schemes that relax these strong assumptions (e.g. Bölöni et al., 2021) may be extended by MSVC in the future. We finally note that the details of the simulated MSVC in a GW-resolving model depend crucially on the subgrid-scale (SGS) diffusion scheme. The HIAMCM employs a Smagorinsky-type diffusion scheme (Becker and Vadas, 2020, Appendix A). Even though this scheme can be considered as a physics-based SGS model and is therefore an improvement compared to adhoc numerical damping methods, it violates the scale-invariance constraint. Scale invariance should be fulfilled by any SGS model when the resolved flow is truncated in a macro-turbulent inertial range (e.g. Schaefer-Rolffs et al., 2014; Schaefer-Rolffs and Becker, 2018, see also references therein). Improved methods should therefore be based on the so-called dynamic Smagorinsky model as discussed in Becker et al. (2023). Corresponding progress in model development and ongoing improvements in computer technology are expected to allow for new insight into MSVC and its relevance to the dynamics in the middle and upper atmosphere.

Acknowledgments: EB was supported by NASA grants 80NSSC22K0174, 80NSSC19K0834, and 80NSSC21M0180. SLV was supported by NSF grant 2329957 and by NASA grant 80NSSC24K0274. The McMurdo lidar project was supported by NSF grant OPP-2110428, and XC was in part supported by AGS-2330168. We are indebted to the editor Ruth Lieberman and to our collaborators from the previous studies that gave rise to this review paper.

References

Achatz, U., 2007a. Gravity-wave breaking: Linear and primary nonlinear dynamics. Adv. Space Res. 40, 719–733. doi:10.1016/j.asr.2007.03.078.

Achatz, U., 2007b. Modal and nonmodal perturbations of monochromatic high-frequency gravity waves: Primary nonlinear dynamics. J. Atmos. Sci. 64, 1977–1994. doi:10.1175/JAS3940.1.

Achatz, U., 2007c. The primary nonlinear dynamics of modal and nonmodal perturbations of monochromatic inertia-gravity waves. J. Atmos. Sci. 64, 74–95. doi:10.1175/JAS3827.1.

Achatz, U., Alexander, M.J., Becker, E., Chun, H.Y., Dörnbrack, A., Holt, L., Plougonven, R., Polichtchouk, I., Sato, K., Sheshadri, A., Stephan, C.C., van Nierkerk, A., Wright, C.J., 2024. Atmospheric gravity waves: Processes and parameterization. J. Atmos. Sci. 81, 237–262. doi:10.1175/JAS-D-23-0210.1.

Avsarkisov, V., Becker, E., Renkwitz, T., 2022. Turbulent coherent structures in the middle atmosphere: Theoretical estimates deduced from a gravity-wave resolving general circulation model. J. Atmos. Sci. doi:org/10.1175/JAS-D-21-0005.1.

Becker, E., 2009. Sensitivity of the upper mesosphere to the Lorenz energy cycle of the troposphere. J. Atmos. Sci. 66, 648–666. doi:10.1175/2008JAS2735.1.

Becker, E., 2012. Dynamical control of the middle atmosphere. Space Sci. Rev. 168, 283–314. doi:10.1007/s11214-011-9841-5.

Becker, E., 2017. Mean-flow effects of thermal tides in the mesosphere and lower thermosphere. J. Atmos. Sci. 74, 2043–2063. doi:10.1175/JAS-D-16-0194.1.

Becker, E., Burkhardt, U., 2007. Nonlinear horizontal diffusion for GCMs. Mon. Wea. Rev. 135, 1439–1454. doi:10.1175/MWR3348.1.

Becker, E., Garcia, R., Pedatella, N., Vadas, S., Yudin, V., 2023. Explicit simulation of gravity waves in whole atmosphere models. Bulletin of the AAS 55. Https://baas.aas.org/pub/2023n3i027.

Becker, E., Goncharenko, L., Harvey, V.L., Vadas, S.L., 2022a. Multi-step vertical coupling during the January 2017 sudden stratospheric warming. J. Geophys. Res. Space Phys. 127. doi:10.1029/2022JA030866.

Becker, E., Grygalashvyly, M., Sonnemann, G.R., 2020. Gravity wave mixing effects on the OH*-layer. Advances in Space Research 65, 175–188. doi:10.1016/j.asr.2019.09.043.

Becker, E., McLandress, C., 2009. Consistent scale interaction of gravity waves in the doppler spread parameterization. J. Atmos. Sci. 66, 1434–1449. doi:10.1175/2008JAS2810.1.

Becker, E., Oberheide, J., 2023. Unexpected DE3 tide in the southern summer mesosphere. Geophys. Res. Lett. 50. doi:10.1029/2023GL104368.

Becker, E., Vadas, S.L., 2018. Secondary gravity waves in the winter mesosphere: Results from a high-resolution global circulation model. J. Geophys. Res. Atmos. 123. doi:10.1002/2017JD027460.

Becker, E., Vadas, S.L., 2020. Explicit global simulation of gravity waves in the thermosphere. J. Geophys. Res. Space Phys. doi:10.1029/2020JA028034.

Becker, E., Vadas, S.L., Bossert, K., Harvey, V.L., Zülicke, C., Hoffmann, L., 2022b. A high-resolution whole-atmosphere model with resolved gravity waves and specified large-scale dynamics in the troposphere and stratosphere. J. Geophys. Res. Atmos. 127. doi:10.1029/2021JD035018.

Bölöni, G., Kim, Y.H., Borchert, S., Achatz, U., 2021. Toward transient subgrid-scale gravity wave pepresentation in atmospheric models. part i: Propagation model including nondissipative wave-mean-flow interactions. J. Atmos. Sci. 78, 1317–1338. doi:10.1175/JAS-D-20-0065.1.

Brune, S., Becker, E., 2013. Indications of stratified turbulence in a mechanistic GCM. J. Atmos. Sci. 70, 231–247. doi: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. J. Atmos. Sol.-Terr. Phys. doi: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). J. Geophys. Res. 118. doi:10.1002/jgrd.50318.

Chen, C., Chu, X., Zhao, J., Roberts, B.R., Yu, Z., Fong, W., Lu, X., Smith, J.A., 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). J. Geophys. Res. Space Physics 121, 1483–1502. doi:10.1002/2015JA022127.

Chu, X., Gardner, C.S., Li, X., Lin, C.Y.T., 2022. Vertical transport of sensible heat and meteoric Na by the complete temporal spectrum of gravity waves in the MLT above McMurdo (77.84°S, 166.69°E), Antarctica. J. Geophys. Res. Atmos. 127. doi:10.1029/2021JD035728.

Chu, X., Yu, Z., Gardner, C.S., Chen, C., Fong, W., 2011. Lidar observations of neutral Fe layers and fast gravity waves in the thermosphere (110-155 km) at McMurdo (77.8°s, 166.7°e), Antarctica. Geophys. Res. Lett. 38. doi:10.1029/2011GL050016.

Chu, X., Zhao, J., Lu, X., Harvey, V.L., Jones, R.M., Becker, E., Chen, C., Fong, W., Yu, Z., Roberts, B.R., Dörnbrack, A., 2018. Lidar observations of stratospheric gravity waves from 2011 to 2015 at McMurdo (77.84°S, 166.69°E), Antarctica: 2. Potential energy densities, lognormal distributions, and seasonal variations. J. Geophys. Res. Atmos. 123. doi:10.1029/2017JD027386.

Crowley, G., Rodrigues, F.S., 2012. Characteristics of traveling ionospheric disturbances observed by the TIDDBIT sounder. Radio Science 47. doi:10.1029/2011RS004959.

Dong, W., Fritts, D.C., Lund, T.S., Wieland, S.A., Zhang, S., 2020. Self-acceleration and instability of gravity-wave packets: 2. Two-dimensional packet propagation, instability dynamics, and transient flow responses. J. Geophys. Res. Atmos. doi:10.1029/2019JD030691.

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. J. Geophys. Res. 107, ACL9–1–ACL9–14. doi:10.1029/2001JD000479.

Forbes, J.M., 2007. Dynamics of the thermosphere. J. Met. Soc. Japan 85B, 193–213.

Frissell, N.A., Baker, J.B.H., Ruohoniemi, J.M., Greenwald1, R.A., Gerrard, 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. J. Geophys. Res. Space Physics 121, 3722–3739. doi:10.1002/2015JA022168.

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

Fritts, D.C., Dong, W., Lund, T.S., Wieland, S.A., 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. J. Geophys. Res. Atmos. doi:10.1029/2019JD030692.

Guo, Y., Liu, A.Z., Gardner, C.S., 2017. First Na lidar measurements of turbulence heat flux, thermal diffusivity, and energy dissipation rate in themesopause region. Geophys. Res. Lett. 44, 5782–5790. doi:10.1002/2017GL073807.

Harding, B.J., Makela, J.J., Englert, C.R., Marr, K.D., Harlander, J.M., England, S.L., Immel, T.J., 2017. The MIGHTI wind retrieval algorithm: Description and verification. Space Sci. Rev. 212, 585–600. doi:10.1007/s11214-017-0359-3.

Harvey, V.L., Pedatella, N., Becker, E., Randall, C.E., 2022. Evaluation of polar winter mesopause wind in WACCMX+DART. J. Geophys. Res. Atmos. doi:10.1029/2022JD037063.

Heale, C.J., Bossert, K., Vadas, S.L., 2022a. 3D numerical simulation of secondary wave generation from mountain wave breaking over Europe. J. Geophys. Res. Atmos. 127. doi:10.1029/2021JD035413.

Heale, C.J., Inchin, P.A., Snively, J.B., 2022b. Primary versus secondary gravity wave responses at F-region heights generated by a convective source. J. Geophys. Res. Atmos. 127. doi:10.1029/2019JD031662.

Hindley, N.P., Mitchell, N.J., Cobbett, N., Smith, A.K., Fritts, D.C., Janches, D., Wright, C.J., Moffat-Griffin, T., 2022. Radar observations of winds, waves and tides in the mesosphere and lower thermosphere over South Georgia Island (54°s, 36°w) and comparison to WACCM simulations. Atmos. Chem. Phys. doi:10.5194/acp-2021-981.

Hines, C.O., 1997. Doppler-spread parameterization of gravity-wave momentum deposition in the middle atmosphere. part 1: Basic formulation. J. Atmos. Sol.-Terr. Phys. 59, 371–386. doi:10.1016/S1364-6826(96)00079-X.

Hoffmann, P., Becker, E., Singer, W., Placke, M., 2010. Seasonal variation of mesospheric waves at northern middle and high latitudes. J. Atmos. Sol.-Terr. Phys. 72, 1068–1079.

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

Huba, J.D., Becker, E., Vadas, S.L., 2023. Simulation study of the 15 january 2022 tonga event: Development of super equatorial plasma bubbles. Geophys. Res. Lett. 50. doi:10.1029/2022GL101185.

Immel, T.J., England, S.L., Mende, B., Heelis, R.A., Englert, C.R., Edelstein, J., Frey, H.U., Korpela, E.J., abd W. W. Craig, E.R.T., Harris, S.E., Bester, M., Bust, G.S., Crowley, G., Forbes, J.M., Gérard, J.C., Harlander, J.M., Huba, J.D., Hubert, B., Kamalabadi, F., Makela, J.J., Maute, A.I., Meier, R.R., Raftery, C., Rochus, P., Siegmund, O.H.W., Stephan, A.W., Swenson, G.R., Frey, S., Hysell, D.L., Saito, A., Rider, K.A., Sirk, M.M., 2018. The Ionospheric Connection Explorer Mission: Mission Goals and Design. Space Sci. Rev. 214. doi:10.1007/s11214-017-0449-2.

Karlsson, B., Becker, E., 2016. How does interhemispheric coupling contribute to cool down the summer polar mesosphere? J. Clim. 29, 8807–8821. doi:10.1175/JCLI-D-16-0231.1.

Knobloch, S., Kaifler, B., Dörnbrack, A., Rapp, M., 2023. Horizontal wavenumber spectra across the middle atmosphere from airborne lidar observations during the 2019 southern hemispheric SSW. Geophys. Res. Lett. 50. doi:10.1029/2023GL104357.

Körnich, K., Becker, E., 2010. A simple model for the interhemispheric coupling of the middle atmosphere circulation. Adv. Space Res. 45, 661–668. doi:10.1016/j.asr.2009.11.001.

Lindborg, E., 2006. The energy cascade in a strongly stratified fluid. J. Fluid Mech. 550, 207–242.

Lindzen, R.S., 1981. Turbulence and stress owing to gravity wave and tidal breakdown. J. Geophys. Res. 86, 9707–9714. doi:10.1029/JC086iC10p09707.

Liu, H.L., Lauritzen, P.H., Vitt, F., 2024a. Impacts of gravity waves on the thermospheric circulation and composition. Geophys. Res. Lett. 51. doi:10.1029/2023GL107453.

Liu, H.L., Lauritzen, P.H., Vitt, F., Goldhaber, S., 2024b. Assessment of gravity waves from tropopause to thermosphere and ionosphere in high-resolution WACCM-X simulations. Journal of Advances in Modeling Earth Systems 16. doi:10.1029/2023MS004024.

Liu, H.L., Marsh, D.R., She, C.Y., Wu, Q., Xu, J., 2009. Momentum balance and gravity wave forcing in the mesosphere and lower thermosphere. Geophys. Res. Lett. 36. doi:10.1029/2009GL037252.

Liu, H.L., McInerney, 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. Geophys. Res. Lett. 41. doi:10.1002/2014GL062468.

Liu, H.L., Wang, W., Huba, J.D., Lauritzen, P.H., Vitt, F., 2023. Atmospheric and ionospheric responses to Hunga-Tonga volcano eruption simulated by WACCM-X. Geophys. Res. Lett. 50. doi:10.1029/2023GL103682.

Lübken, F.J., 1997. Seasonal variation of turbulent energy dissipation rates at high latitudes as determined by in situ measurements of neutral density fluctuations. J. Geophys. Res. 102, 13441–13456.

McFarlane, N.A., 1987. The effect of orographically excited gravity wave drag on the general circulation of the lower stratosphere and troposphere. J. Atmos. Sci. 44, 1775–1800.

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. J. Geophys. Res. 111. doi:10.1029/2005JD006776.

Medvedev, A.S., Klaassen, G.P., 2000. Parameterization of gravity wave momentum deposition based on nonlinear wave interactions: Basic formulation and sensitivity tests. J. Atmos. Sol.-Terr. Phys. 62, 1015–1033.

Nayak, C., Yiğit, E., 2019. Variation of small-scale gravity wave activity in the ionosphere during the major sudden stratospheric warming event of 2009. J. Geophys. Res. Space Physics 124, 470–488. doi:10.1029/2018JA026048.

Plougonven, R., Zhang, F., 2014. Internal gravity waves from atmospheric jets and fronts. Rev. Geophys. 52, 33–76. doi:10.1002/2012RG000419.

Podglajen, A., Hertzog, A., Plougonven, R., Legras, B., 2016. Lagrangian temperature and vertical velocity fluctuations due to gravity waves in the lower stratosphere. Geophys. Res. Lett. 43, 3543–3553. doi:10.1002/2016GL068148.

Sato, K., Kohma, M., Tsutsumi, M., Sato, T., 2017. Frequency spectra and vertical profiles of wind fluctuations in the summer Antarctic mesosphere revealed by MST radar observations. J. Geophys. Res. Atmos. 122, 3–19. doi:10.1002/2016JD025834.

Sato, K., Tanteno, S., Watanabe, S., Kawatani, Y., 2012. Gravity wave characteristics in the southern hemisphere revealed by a high-resolution middle-atmosphere general circulation model. J. Atmos. Sci. 69, 1378–1396. doi:10.1175/JAS-D-11-0101.1.

Sato, K., Tsuchiya, C., Alexander, M.J., Hoffmann, L., 2016. Climatology and ENSO-related interannual variability of gravity waves in the Southern Hemisphere subtropical stratosphere revealed by high-resolution AIRS observations. J. Geophys. Res. Atmos. 121, 7622–7640. doi:10.1002/2015JD024462.

Sato, K., Yoshiki, M., 2008. Gravity wave generation around the polar vortex in the stratosphere revealed by 3-hourly radiosonde observations at Syowa Station. J. Atmos. Sci. 65, 3719–3735. doi:10.1175/2008JAS2539.1.

Schaefer-Rolffs, U., Becker, E., 2018. Scale-invariant formulation of momentum diffusion for high-resolution atmospheric circulation models. Mon. Wea. Rev. 146, 1045–1062. doi:10.1175/MWR-D-17-0216.1.

Schaefer-Rolffs, U., Knöpfel, R., Becker, E., 2014. A scale invariance criterion for les parametrizations. Meteorol. Z. doi:10.1127/metz/2014/0623.

Senf, F., Achatz, U., 2011. On the impact of middle-atmosphere thermal tides on the propagation and dissipation of gravity waves. J. Geophys. Res. 116. doi:10.1029/2011JD015794.

Simmons, A.J., Burridge, D.M., 1981. An energy and angular momentum conserving vertical finite-difference scheme and hybrid vertical coordinates. Mon. Wea. Rev. 109, 758–766.

Smith, A.K., 2012. Global dynamics of the MLT. Surv. Geophys. 33, 1177–1230. doi:10.1007/s10712-012-9196-9.

Smith, S.A., Fritts, D.C., VanZandt, T.E., 1987. Evidence for a saturated spectrum of atmospheric graavity waves. J. Atmos. Sci. 44, 1401–1410. doi:10.1175/JAS-D-15-0324.1.

Themens, D.R., Watson, C., Žagar, N., Vasylkevych, S., Elvidge, S., McCaffrey, A., Prikryl, P., Reid, B., Wood, A., Jayachandran, P.T., 2022. Global propagation of ionospheric disturbances associated with the 2022 Tonga volcanic eruption. Geophys. Res. Lett. 49. doi:10.1029/2022GL098158.

Torr, M.R., Richards, P.G., Torr, D.G., 1981. Solar EUV energy budget of the thermosphere. Adv. Space Res. 1, 53–61. doi:10.1016/0273-1177(81)90417-8.

Vadas, S.L., 2007. Horizontal and vertical propagation and dissipation of gravity waves in the thermosphere from lower atmospheric and thermospheric sources. J. Geophys. Res. 112. doi:10.1029/2006JA011845.

Vadas, S.L., 2013. Compressible f-plane solutions to body forces, heatings, and coolings, and application to the primary and secondary gravity waves generated by a deep convective plume. J. Geophys. Res. Space Physics 118, 2377–2397. doi:10.1002/jgra.50163.

Vadas, S.L., Azeem, I., 2021. Concentric secondary gravity waves in the thermosphere and ionosphere over the continental United States on 25-26 March 2015 from deep convection. J. Geophys. Res. Space Phys. doi:10.1029/2020JA028275.

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. J. Geophys. Res. Atmos. 123, 9326–9369. doi: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. J. Geophys. Res. Space Phys. 124, 7687–7718. doi:10.1029/2019JA026694.

Vadas, S.L., Becker, E., Bossert, K., Baumgarten, G., Hoffmann, L., Harvey, V.L., 2023a. Secondary gravity waves from the stratospheric polar vortex over ALOMAR observatory on 12-14 January 2016: Observations and modeling. J. Geophys. Res. Atmos. 128. doi:10.1029/2022JD036985.

Vadas, S.L., Becker, E., Bossert, K., Hozumi, Y., Stober, G., Harvey, V.L., Baumgarten, G., Hoffmann, L., 2024. The role of the polar vortex jet in the generation of primary and higher-order gravity waves in the stratosphere, mesosphere and thermosphere during 11-14 january 2016. J. Geophys. Res. Space Phys. 129. doi:10.1029/2024JA032521.

Vadas, S.L., Becker, E., Figueiredo, C.A.O.B., Bossert, K., Harding, B.J., Gasque, L.C., 2023b. Primary and secondary gravity waves and large-scale wind changes generated by the Tonga volcanic eruption on 15 January 2022: Modeling and comparison with ICON-MIGHTI winds. J. Geophys. Res. Space Phys. 128. doi:10.1029/2022JA031138.

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. J. Geophys. Res. Space Physics 115. doi:10.1029/2009JA015053.

Vadas, S.L., Figueiredo, C.A.O.B., Becker, E., Huba, J.D., Themens, D.R., Hindley, N.P., Galkin, I., Bossert, K., 2023c. Traveling ionospheric disturbances induced by the secondary gravity waves from the tonga eruption on 15 January 2022: Modeling with MESORAC/HIAMCM/SAMI3 and comparison with GPS/TEC and ionosonde data. J. Geophys. Res. Space Phys. 128. doi:10.1029/2023JA031408.

Vadas, S.L., Fritts, D.C., 2001. Gravity wave radiation and mean responses to local body forces in the atmosphere. J. Atmos. Sci. 58, 2249–2279.

Vadas, S.L., Fritts, D.C., 2002. The importance of spatial variability in the generation of secondary gravity waves from local body forces. Geophys. Res. Lett. 29. doi:10.1029/2002GL015574.

Vadas, S.L., Fritts, D.C., Alexander, M.J., 2003. Mechanisms for the generation of secondary waves in wave breaking regions. J. Atmos. Sci. 60, 194–214. doi:10.1175/1520-0469(2003)060¡0194:MFTGOS¿2.0.CO;2.

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. J. Geophys. Res. 114. doi: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. J. Geophys. Res. 118, 2593–2617. doi: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. J. Geophys. Res. 119. doi:10.1002/2014JA020280.

Vadas, S.L., Zhao, J., Chu, X., Becker, E., 2018. The excitation of secondary gravity waves from local body forces: Theory and observation. J. Geophys. Res. Atmos. 123, 9296–9325. doi:10.1029/2017JD027970.

Xu, S., Vadas, S.L., Yue, J., 2021. Thermospheric traveling atmospheric disturbances in austral winter from GOCE and CHAMP. J. Geophys. Res. Space Phys. 126. doi:10.1029/2021JA029335.

Zhang, S.R., Vierinen, J., Aa, E., Goncharenko, L.P., Erickson, P., Rideout, W., Coster1, A.J., Spicher, A., 2022. Tonga volcanic eruption induced global propagation of ionospheric disturbances via Lamb waves. Frontiers in Astronomy and Space Sciences 9. doi:10.3389/fspas.2022.871275.

Zhao, J., Chu, X., Chen, C., Lu, X., Fong, W., Yu, Z., Jones, R.M., Roberts, B.R., Dörnbrack, A., 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. J. Geophys. Res. Atmos. 122. doi:10.1002/2016JD026368.