

Journal of Atmospheric and Solar-Terrestrial Physics 68 (2006) 247-265



www.elsevier.com/locate/jastp

Mean and variable forcing of the middle atmosphere by gravity waves

David C. Fritts*, Sharon L. Vadas, Kam Wan, Joseph A. Werne

Colorado Research Associates, a Division of NorthWest Research Associates, 3380 Mitchell Lane, Boulder, CO 80301, USA

Available online 2 November 2005

Abstract

Recently we have begun to appreciate more fully the degree and the consequences of variability in gravity wave (GW) forcing of the middle atmosphere. Such variability arises for a number of reasons. GW sources in the lower atmosphere reflect the significant spatial and temporal variability of normal meteorological processes. GW amplitudes and characteristics are modulated by the wind and temperature fields through which they propagate. Nonlinear interactions and instability appears to be greatest among GWs occurring at the smaller spatial scales and periods that account for the majority of energy and momentum transports into the middle atmosphere. This paper both surveys recent findings and introduces new results.

© 2005 Elsevier Ltd. All rights reserved.

Keywords: Gravity waves; Middle atmosphere dynamics; Instability dynamics; Turbulence

1. Introduction

We have known for many years that gravity wave (GW) amplitudes and characteristics exhibit considerable variability in the lower and middle atmosphere. The dominant sources in the lower atmosphere, including convection, orography, and wind shear, all yield GWs having spatial and temporal scales reflecting source characteristics because GW excitation can often be viewed as a linear or quasi-linear response to what is often a highly nonlinear process. The various sources of GWs in the middle atmosphere, and the GW characteristics arising from them, were recently

*Corresponding author. Tel.: +13034159701x205; fax: +13034159702.

E-mail address: dave@co-ra.com (D.C. Fritts).

reviewed in some detail by Fritts and Alexander (2003). It is generally believed, for example, that both the spatial scales and temporal behavior of convection and the environmental shears in which convection occurs contribute to the characteristics and anisotropy of the resulting GW field. Orography also leads most often to a linear GW response that reflects the spatial scales and orientation of the terrain. Wind shear, in contrast, may favor excitation via a nonlinear mechanism, so-called "envelope" radiation, because the linear growth rates of the larger horizontal scales that can propagate vertically away from the shear are much smaller than the growth rates of the smaller-scale Kelvin-Helmholtz (KH) instability that itself imposes the envelope (or packet) scale. Other source mechanisms are also operative, including frontal dynamics and adjustment processes, and contribute to the

^{1364-6826/\$ -} see front matter 2005 Elsevier Ltd. All rights reserved. doi:10.1016/j.jastp.2005.04.010

GW spectrum over a wide range of scales, but these have not been studied as extensively to date (see Fritts and Alexander, 2003).

Once excited, GWs propagate through an atmosphere that itself exhibits variability on all possible spatial and temporal scales. Over a large range of scales, $\sim 1-1000$ km and a few minutes to tens of hours, this variability is imposed by the GW spectrum itself. Variability at smaller and larger scales accompanies turbulence arising most often from GW instability processes (and convection at lower altitudes) and planetary-scale motions (tides and planetary waves). A variable environment contributes both coherent (systematic) and incoherent (random) variations in GW properties, with the coherent responses determined to a large extent by a linear dispersion relation (and ray equations) relating GW phase speeds (or intrinsic frequencies) and wavenumbers to environmental density, stratification, winds, and shears. GW scales range from horizontal wavelengths of $\sim 10-1000$ km and vertical wavelengths of $\sim 1-100$ km and intrinsic frequencies range from the inertial to the buoyancy frequency, with both the vertical wavelength and the intrinsic frequency determined initially by source conditions and thereafter by the GW propagation environment. Of greatest relevance in this paper are the variations that occur in intrinsic phase speed (or vertical wavenumber) and vertical group velocity, since GWs attaining large amplitudes and high vertical group velocities can dominate GW influences at greater altitudes.

Nonlinear dynamics occurring within a GW field also contribute to, and may indeed amplify, variability arising from linear dynamics alone. This is because all GWs are inherently unstable to a wide range of perturbations. At smaller amplitudes, these instabilities manifest as systematic exchanges of energy among GWs exhibiting weak "resonant" interactions (Klostermeyer, 1991; Vanneste, 1995; Sonmor and Klaassen, 1997) that are believed not to impact GW momentum transport. At larger amplitudes, however, these instabilities lead to GW dissipation, breaking, turbulence, spectral energy transfers, and divergent momentum fluxes that play prominent roles in middle atmosphere dynamics (Fritts, 1984a, 1989; Dunkerton, 1987, 1989; Sonmor and Klaassen, 1997; Fritts and Alexander, 2003). Instability dynamics also impose both mean and spatially and temporally localized flux divergence that forces the zonal mean circulation and excites additional GWs at higher altitudes (Fritts

et al., 2003; Vadas and Fritts, 2001; Vadas et al., 2003).

The purpose of this paper is to review and update our understanding of the causes and effects of mean and variable GW forcing of the middle atmosphere. Characteristics and influences arising from various GW sources are described in Section 2. These suggest very different influences by different sources, both in GW character and in their geographic and temporal distributions. Propagation effects and contributions to forcing variability are discussed in Section 3. These provide several mechanisms by which localized GW forcing can arise. Section 4 reviews the implications of nonlinear interaction and instability processes for mean and variable GW forcing. Nonlinear processes are likely the least quantified at this time: we are, nevertheless. beginning to understand the range of important dynamics in a qualitative manner. Our results are summarized in Section 5.

2. GW source characteristics

This section reviews GW character and variability due to the dominant GW sources as we understand them at present. As noted above, the sources we believe to dominate GW excitation in the lower atmosphere are convection, orography, and wind shear, though these vary with geography, season, and local meteorology. Other sources are also important under certain conditions and/or for specific portions of the GW spectrum. These include frontal dynamics, adjustment of unbalanced flows, wave–wave interactions, local body forcing (effectively the local acceleration accompanying GW dissipation and momentum flux divergence), and a few others likely less important.

2.1. Convection

Both satellite and balloon studies have revealed that GW variances are largest throughout the stratosphere at equatorial latitudes (Fetzer and Gille, 1994, 1996; Allen and Vincent, 1995; Tsuda et al., 2000; Alexander et al., 2000). This appears to be largely a result of the importance of deep convection as a major source of GWs and the presence of most deep convection in the tropics (Salby and Garcia, 1987; Taylor and Hapgood, 1988; Fritts and Nastrom, 1992). It has also been argued, however, that the large mean variances at tropical latitudes are due, in part at least, to an observational bias favoring the detection of slow moving inertia-GWs (Alexander et al., 2002).

Deep, fast convection excites GW having large vertical scales and high intrinsic frequencies and phase speeds (Fovell et al., 1992; Dewan et al., 1998; Piani et al., 2000; Lane et al., 2001; Horinouchi et al., 2002; Sentman et al., 2003; Vadas and Fritts, 2004). Mesoscale convective complexes (MCC) also organize convection on much larger scales, however, and additionally excite considerable GW activity near inertial frequencies that may have significant influences to high altitudes (Pfister et al., 1986; Tsuda et al., 1994; Karoly et al., 1996; Shimizu and Tsuda, 1997; Garcia and Sassi, 1999; Wada et al., 1999; Vincent and Alexander, 2000).

Three mechanisms are thought to describe approximately GW generation by convection. These are (1) thermal forcing via latent heat release that excites vertical scales comparable to the forcing depth (Bergman and Salby, 1994; Alexander et al., 1995; Piani et al., 2000), (2) the "obstacle" or "transient mountain" effect in which wind shear at cloud top imposes relative motion over convective cells (Clark et al., 1986; Pfister et al., 1993a, b; Alexander and Vincent, 2000: Vincent and Alexander, 2000), and (3) a "mechanical oscillator" effect in which oscillatory convective plumes project those periods onto the GW field (Fovell et al., 1992; Lane et al., 2001). More recently, Lane and Clark (2002) revisited the convective boundary layer problem and concluded that GWs were excited primarily by oscillatory motions with GW structures determined largely by filtering thereafter. The relative importance of heat and momentum flux convergence and diabatic forcing was also examined by Song et al. (2003). They concluded that forcing by net flux convergence is comparable to diabatic heating, that these sources are largely out of phase, and that the net effect is less efficient excitation of GWs that are able to reach the stratosphere.

Other processes also operate to make convection a highly variable GW source. Wind shear in the troposphere yields tilted convection and GWs having that same preferred phase tilt and direction of propagation. While GW scales depend in large part on the spatial scales of convection, the temporal behavior poses an additional constraint, essentially eliminating GW periods shorter than the characteristic time scale of the convection (Vadas and Fritts, 2004). The net effect is convectively generated GWs that span horizontal scales of \sim 10–1000 km and periods of minutes to 10's of hours, with the largest amplitudes, frequencies, and momentum fluxes accompanying convection that is deep, spatially localized, and fast (Piani et al., 2000; Lane et al., 2001; Vadas and Fritts, 2004).

Examples of the GW patterns arising from an analytic description of convective plumes and the corresponding frequency and vertical wavenumber spectrum exhibiting a range of spatial and temporal scales are shown in Fig. 1 (Vadas and Fritts, 2004). For these choices of source scales, the dominant responses and momentum fluxes occur at horizontal and vertical wavelengths of ~ 40 and 14 km and periods of ~15 min (~3 buoyancy periods, $T_{\rm b}$). The forcing geometry and the character of the GWs excited are shown in the left panel of Fig. 2. Deep and relatively narrow sources lead to GWs having large vertical scales and steep phase slopes. Because such motions also have high phase speeds and vertical group velocities, they may penetrate to high altitudes in a variety of environments. It is important to note, however, that those GWs that achieve very high altitudes have similar frequencies $(\omega \sim N/3)$ but significantly larger spatial scales, horizontal and vertical wavelengths of ~ 100 and 40 km, respectively (see Vadas and Fritts, 2004).

2.2. Orography

GW generation by orography has been studied extensively because such waves have effects throughout the atmosphere. At lower altitudes, mountain waves may induce strong local flows and wave drag that influence surface flows and tropospheric jet structure and are important for numerical modeling of tropospheric weather. But GWs generated by orography may also penetrate to much greater altitudes and influence the local and zonal mean structures of the stratosphere, mesosphere, and lower thermosphere (Preusse et al., 2002; Jiang et al., 2002, 2003, 2004; Kim et al., 2003; Fritts and Alexander, 2003). Mountain wave forcing is often approximately linear, though strong nonlinear, or resonant, responses (downslope or "chinook" winds) may develop on the lee slope when atmospheric structure is suitable. Such forcing also varies strongly with terrain height, scale, and orientation. Flow may be around rather than over terrain if the Froude number, Fr = U/Nh, is small or the terrain is three-dimensional (3D) or aligned along the flow; the flow may also separate from the terrain, leading to much smaller GW responses. Like convection, orography excites GWs having a



Fig. 1. GW patterns in vertical velocity and perturbation temperature (left and center cross sections) at 90 km and vertical velocity from 90 to 150 km (right cross sections) above a simulated mesoscale convective complex (MCC). The corresponding distribution of GW energy with vertical wavelength and intrinsic frequency arising from a cluster of convective plumes (right panel) with mean widths and depths of \sim 5–10 km and \sim 1.5–6 km, respectively (see Vadas and Fritts, 2004, for further details).



Fig. 2. Body force geometries and associated GW ray paths for a deep, narrow vertical forcing representative of convective plumes (left) and a shallow, horizontally extensive horizontal forcing more typical of a slow adjustment process (right). In each case, the spatial scales largely determine the frequencies of the radiated GWs, except that the forcing time scale may further restrict the excitation of high frequencies.

wide range of spatial scales and intrinsic frequencies, with the dominant fluxes and effects accompanying waves having large amplitudes, small scales, and high intrinsic frequencies. But unlike convection, mountain waves have phase speeds near zero and thus penetrate to high altitudes only when sufficiently nonzero winds (along the plane of propagation) extend throughout the atmospheric column, a process that appears to occur at only a few preferred sites (Jiang et al., 2002; Preusse et al., 2002). While 3D orography induces 3D GW propagation (Broutman et al., 2002, 2003, 2004), the dominant momentum fluxes are associated with GWs propagating upstream relative to the local mean flow exciting the GW response.

2.3. Wind shear

Wind shear is believed to be a statistically significant source of GWs near the tropopause and at higher altitudes, based on modeling and observational studies. Modeling has suggested that the most likely source mechanism may be envelope radiation, as linear growth of larger-scale GWs cannot compete with that of the KH instability (Fritts, 1984b; Chimonas and Grant, 1984; Scinocca and Ford, 2000; Bühler et al., 1999). Other studies have revealed enhanced variances in the vicinity of jet stream shears or argued that such a source is likely to contribute significantly to the momentum budget of the middle atmosphere (Fritts and Nastrom, 1992; Bühler and McIntyre, 1999). Thus, while we believe that wind shear is likely an important source, this source is more poorly understood and quantified than convection and orography at this time. We do know, however, that phase speeds must be comparable to mean winds at the source altitude (from the Miles-Howard semi-circle theorem). We suspect that such forcing leads to horizontal scales that are 10's rather than hundreds or thousands of km because KH patches appear nearly always to be limited in their horizontal extent (Fritts and Alexander, 2003; Hecht et al., 2005). Finally, we know that such a nonlinear process is highly intermittent because it results from the exponential growth of KH shear instability, in which KH growth and turbulent breakdown occupies only a few T_b (Palmer et al., 1996; Fritts et al., 1996a; Werne and Fritts, 1999).

2.4. Adjustment processes

Adjustment processes of unbalanced flows encompass a wide range of approximations and dynamics (McIntyre, 2003). The most simplistic of these is an imbalance at some order of approximation that arises either from an initial unbalanced state (Rossby adjustment) or as the large-scale flow evolves from (or is perturbed from) a balanced to an unbalanced state ("spontaneous" adjustment) (see McIntyre, 2003). In general problems, this adjustment involves alteration of the 3D wind and geopotential fields to attain a new balanced state and the radiation of inertia-GWs (IGWs) to accommodate energy and momentum conservation. Spatial and temporal scales for such processes vary widely, with spontaneous adjustment of large-scale flows (such as jet structures or troposphere-stratosphere exchange events) occurring on scales of hundreds or thousands of km and many hours. At the other end of the spectrum, instability dynamics, specifically shear instability and GW breaking, can result in small-scale flows (10's of km or less) that evolve to an unbalanced state on time scales of an hour or less.

Large-scale adjustment processes having long time scales are illustrated schematically in the right panel of Fig. 2, with emergence of IGWs having large horizontal scales, much smaller vertical scales (because jet streams are much thinner than they are wide), and a dominance of intrinsic frequencies near the inertial frequency (Fritts and Luo, 1992; Luo and Fritts, 1993; Vadas and Fritts, 2001). Smallerscale adjustment processes can exhibit a wide range of spatial and temporal scales, depending on the geometry and the time scale of the event. For example, rapid events can be triggered by rearrangement of the local wind and temperature structure due to KH shear instability (Bühler et al., 1999) occurring on spatial scales of $\sim 1-10$'s of km and time scales of a few $T_{\rm b}$ (Werne and Fritts, 1999; Fritts and Alexander, 2003) or to the body forcing (see below) accompanying GW momentum flux divergence in a local breaking event. In all cases, however, the resulting GW scales are determined by a combination of event spatial and temporal scales.

A deep, narrow, fast event will lead to GW excitation resembling the left panel in Fig. 2. A wide, shallow event (right panel of Fig. 2) will lead to IGW excitation only, independent of whether the time scale is fast or slow, because the source has no spatial components having steep phase slopes. In cases where the event spatial scales are deep and narrow, but the time scales are long, there is a mismatch between the intrinsic frequencies implied by the spatial geometry and the slow evolution of the flow, and the radiation of high-frequency GWs is suppressed (Lighthill, 1978, Section 4.9; Vadas and Fritts, 2001). Indeed, as the time scale for adjustment becomes very long. GW radiation becomes negligible, but the new balanced mean state is independent of the time scale of the adjustment (Vadas and Fritts, 2001; Bühler and McIntyre, 2005). The IGW field and two-dimensional (2D) wavenumber-frequency spectrum arising from spontaneous adjustment having Gaussian geometry, a length, width, and depth of 500, 100, and 2 km (full-width, half-maximum, FWHM), respectively, and a time scale of 1h (FWHM) is shown for comparison with the response to deep, fast forcing in Fig. 3. The results displayed here employed a Fourier-Laplace transform that represents an exact solution of the linear Boussinesq equations (Vadas and Fritts, 2001).

2.5. Local body forces

The role of localized GW breaking as a source of middle atmosphere variability and additional GWs having significant influences at higher altitudes is likely under-appreciated at present. However, the tendencies (1) for sources of high-frequency GWs to be spatially localized, intermittent, and strong, (2) for GWs having large-amplitudes and large momentum fluxes to also be spatially localized at higher altitudes, and (3) for instability dynamics to lead to rapid, local breaking and momentum flux



Fig. 3. GW radiation resulting from spontaneous adjustment of a zonal jet centered at 10 km and having Gaussian cross sections with FWHM dimensions of 500, 100, and 2 km in (x, y, z), respectively, and a time scale of 1 h. The right panel shows the 2D vertical wavenumber-frequency spectrum of these motions. Note the predominance of GWs having intrinsic frequencies of $\sim N/50$, corresponding to the scales imposed by the forcing geometry.

divergence (see below) together suggest that GW forcing of the large-scale flow is more likely to be intermittent and variable than smooth, systematic, and uniform. Thus, while the mean response to GW forcing is the same whether forcing is uniform or highly variable in time (apart, perhaps, from influences of larger-scale wave motions and filtering, see below), local GW instability, momentum flux divergence, and forcing are likely an increasingly important source of additional GWs at higher altitudes.

GWs arise from wave breaking regions in two ways. One is via direct nonlinear interactions that excite other GWs at smaller scales (or higher harmonics) of the parent GW (see below). The second is via the adjustment process accompanying rapid, local body forcing due to local instability and momentum flux divergence and has been addressed by Zhu and Holton (1987), Vadas and Fritts (2002), and Vadas et al. (2003). As discussed above, the spatial scales and intrinsic frequencies of the resulting GWs depend on the spatial and temporal scales of the body forcing event, but it is noteworthy that deep, rapid instability processes can lead to GWs having significantly larger scales and vertical group velocities than the GW undergoing instability. In such cases, the radiated GWs may penetrate to, and have influences at, very much higher altitudes than the initial GW itself. One additional caveat that is important to note is the requirement that the time scale for initial GW propagation, essentially its period, be shorter than the time scale of the radiated GWs, for the same reasons as discussed for adjustment processes above (Bühler and McIntyre, 2005).

An example of an apparent strong local GW breaking and body forcing event seen in the OH airglow emission by Yamada et al. (2001) and diagnosed by Fritts et al. (2002) is shown in Fig. 4. This event occurred on horizontal and vertical scales of ~50 and 10–20 km, a time scale of a few $T_{\rm b}$ (~10 min or less), and accompanied a GW having a high intrinsic frequency, a large vertical wavelength, and attaining a very large amplitude prior to instability. The estimated spectrum of radiated GWs was found to be qualitatively like that shown in Fig. 1, seeming to confirm the potential importance of such instability events as another source of GWs at higher altitudes.

2.6. Other sources

Additional GW sources likely to be important in the lower and middle atmosphere include frontal dynamics and wave-wave interactions. Others likely play a smaller role. Frontogenesis leads to IGW excitation because of the large horizontal scales involved (Griffiths and Reeder, 1996; Reeder and

1995 223 1628 UT

1995 223 1640 UT

1995 223 1652 UT



Fig. 4. GW structure in OH airglow showing wave amplification and instability. Successive images are spaced by 4 and 12 min, intensity perturbations achieved values as large as 55%, and temperature perturbations were $\sim 10\%$ prior to breaking (after Yamada et al., 2001). The GW horizontal wavelength was ~ 27 km and the spatial scale of the breaking region was inferred to be ~ 50 km horizontally and $\sim 10-20$ km vertically (Fritts et al., 2002).

Griffiths, 1996) and is very like jet stream adjustment as a source. Associated convection and instability dynamics may excite GWs at the other (higher) ends of the wavenumber and frequency spectra, as noted above. Wave-wave interactions operate across the full range of the GW spectrum and will be discussed further below.

2.7. GW energy and momentum flux spectra

Collectively, GW sources and interactions in the lower and middle atmosphere contribute to the establishment of a "mean" GW spectrum having near-universal shapes in frequency and wavenumber, despite many reasons to expect otherwise (VanZandt, 1982; Fritts and VanZandt, 1987, 1993; Tsuda et al., 1991; Nastrom et al., 1997; Fritts and Alexander, 2003). A schematic of the mean spectra of horizontal and vertical GW energy density with intrinsic frequency is shown in the left panel of Fig. 5 and emphasizes what we see in essentially all ground-based observations: GW energy density peaks near the inertial frequency and has a near-universal slope of $\sim -5/3$. Limited in situ measurements of the intrinsic frequency spectrum using constant-pressure balloons by Hertzog and Vial (2001), however, suggest a somewhat steeper slope of the intrinsic frequency spectrum of ~ -2 . The corresponding frequency distribution of momentum flux inferred from the energy spectra and the dispersion relation (and confirmed by multiple observations, see Fritts and Alexander,



Fig. 5. Schematic of GW energy in horizontal and vertical motions (left) and the corresponding momentum flux (right) as functions of intrinsic frequency. Note that at middle and high latitudes, momentum flux is strongly concentrated at high intrinsic frequencies.

2003) is shown in the right panel of Fig. 5. Note that the momentum flux, representing both vertical transports of horizontal momentum and meridional transports of heat (relevant only for IGWs), may be written in the form

$$\overline{u'w'}(1-f^2/\omega^2),$$
 (1)

so that the two contributions exactly cancel at the inertial frequency (Andrews and McIntyre, 1976; Fritts and Alexander, 2003).

These energy and momentum flux spectra emphasize an important point that is often overlooked in assessing GW influences in the lower and middle atmosphere: dominant energies and momentum fluxes occur at *opposite ends* of the frequency spectrum. Thus the GWs having the largest amplitudes, energy densities, and horizontal scales (hence the most easily observable in many data sets) are *not* the GWs having the largest energy and momentum fluxes and atmospheric effects.

3. Influences of propagation

It is well known that GWs are strongly influenced by the environments through which they propagate. In most cases, this propagation can be described by approximately linear dynamics. There are, however, circumstances, especially for larger-amplitude GWs or for GWs in sheared and time-dependent environments, where numerical studies suggest either significant departures from the expectations of linear theory or unusual behavior, GW responses, or structure (Broutman, 1986; Broutman and Young, 1986; Zhong et al., 1995; Broutman et al., 1997; Buckley et al., 1999; Sutherland, 1999, 2000, 2001; Sonmor and Klaassen, 2000; Fritts and Alexander, 2003). Representative ray paths for GWs having varying initial intrinsic frequencies in a sheared environment are shown in Fig. 6. In this example, the mean wind profile exhibits a westward shear above the GW source which refracts GWs having both eastward and westward phase velocities. To understand linear GW propagation, we employ the approximate linear dispersion relation for GWs propagating in an east-west plane for which rotational and shear effects are not



Fig. 6. Schematic showing GW propagation with ray paths from a localized source. Dashed lines show GWs that refract in the mean shear but maintain finite m^2 and propagate to higher altitudes before dissipating. Dotted lines show GWs that are excluded from higher altitudes due to a critical level encounter (left) and a turning level (right).

important, which is given by

$$m^{2} = k^{2} (N^{2} / \omega^{2} - 1) - 1 / 4H^{2}, \qquad (2)$$

where k and m are the horizontal and vertical wavenumbers, N the buoyancy frequency, $\omega = k(c - U)$ is the intrinsic frequency, c and U are the GW phase speed and mean wind in the direction of GW propagation, and H the density scale height, typically ~7 km.

GWs having eastward propagation experience increasing ω within the wind shear. Initial frequencies that cause N^2/ω^2 to become sufficiently small that m^2 falls below zero within the shear layer become evanescent at greater altitudes (they encounter a turning level) and reflect in the vertical. Other GWs having larger initial *m* and shallower propagation paths also exhibit refraction to smaller *m* and higher intrinsic frequencies (steeper propagation angles), but maintain positive m^2 , continue their upward propagation, and induce an eastward body force where these GWs are dissipated.

GWs having westward propagation experience decreasing ω within the shear layer. In this case, initial frequencies for which the corresponding phase speed, c, is less negative than the maximum negative wind speed, ω becomes zero (where c = U) and the linear dispersion relation implies that the GW encounters a critical level and is trapped at this level. For GWs having larger negative phase speeds, ω remains finite, c < U everywhere, and these GWs refract to smaller vertical scales and shallower propagation, but retain their upward propagation and apply a westward body force where they dissipate.

Departures from this simple picture arise for a number of reasons. Two of those that impact GW spectral evolution are time dependence and a component of vertical motion of the local mean flow. Together, these result in significantly different interactions among diverse scales of motions, and different implications for spectral energy transfers, than when these effects are neglected (Bruhwiler and Kaper, 1995; Zhong et al., 1995; Broutman et al., 1997; Eckermann, 1997; Walterscheid, 2000). GW transience and packet localization likewise have some interesting effects, among them GW instability accompanying "self-acceleration" and instability and permanent mean-flow changes at large GW amplitudes during turning level encounters (Sutherland, 1999, 2000, 2001).

Finally, GWs are strongly modulated by tidal and planetary wave motions, leading to strong GW

filtering (Walterscheid, 1981; Smith, 1996) and significant modulations of GW variances and momentum fluxes at tidal (Fritts and Vincent, 1987; Wang and Fritts, 1991) and planetary wave periods (Thayaparan et al., 1995; Isler and Fritts, 1996; Nakamura et al., 1997; Manson et al., 1998). These GW modulations lead, in turn, to feedbacks on tidal and planetary wave structures, but there is considerable uncertainty at this time, with the magnitude (and sign) of the effect dependent on the GW parameterization employed (see Fritts and Alexander, 2003, for a review).

4. Nonlinear processes

For the discussion here, we classify nonlinear processes as wave-wave interactions, instability dynamics, or wave-mean flow interactions, though, as will be seen below, this is somewhat oversimplified and, depending on the scales involved, a number of processes can be viewed from more than one perspective. We will consider wave-wave interactions to include dynamics that can be described approximately by weakly nonlinear resonant interactions among three GWs (or more generally among GWs and vortical modes, see Mied, 1976; McComas and Bretherton, 1977; Yeh and Liu, 1981; Müller et al., 1986; Dong and Yeh, 1988; Yeh and Dong, 1989; Dunkerton, 1989; Fritts and Alexander, 2003). Instability dynamics, in our discussion, include smaller-scale processes that are typically 3D, comprise "tube-like" vortex structures, occur within a preferred phase of a largeamplitude GW, and arise by extracting energy from the GW through buoyancy and/or shear sources (Klostermeyer, 1991; Lombard and Riley, 1996; Fritts et al., 1996b, 1998, 2003; Sonmor and Klaassen, 1997; Fritts and Alexander, 2003). Wave-mean flow interactions include the responses of the mean flow (or larger-scale motions) to GW momentum flux divergence and the mean and spatially and temporally localized body forces that arise from this divergence (Vincent and Reid, 1983; Fritts and Vincent, 1987; Zhu and Holton, 1987; Vadas and Fritts, 2001; Fritts et al., 2002). Each of these processes is discussed in greater detail below.

As an example of the complexity that can arise in instability studies, we show in Fig. 7 predictions of instability growth rates for various streamwise and spanwise wavenumbers (α , β) scaled by the GW total wavenumber following the methodology of Lombard and Riley (1996). Here, "streamwise" refers to

the direction of propagation of the GW, whereas "spanwise" refers to the cross-stream, or horizontal orthogonal, direction. These plots reveal that instability alignment, scales, and growth rates are all strong functions of GW amplitude, frequency, and Reynolds number and that there are, in general, multiple possible instability structures having comparable growth rates for any combination of GW parameters. This complicates both applications of instability theory and interpretations of apparent instability structures in observed flows in the atmosphere and the laboratory. It also means that there may be competing instabilities (having very similar growth rates, but very different orientations, scales, and energetic sources) in any specific flow. Similar results are obtained with the methodology of Sonmor and Klaassen (1997), though their presentation emphasizes only those modes having the dominant growth rate at each point in parameter space. Klostermeyer (1991), Lombard and Riley (1996), and Sonmor and Klaassen (1997) all also identify links between instability structures at small and large amplitudes and other modes identified by previous authors (Hines, 1971, 1988; Yeh and Liu, 1981).

The situation is further complicated in recognizing that optimal perturbation theory (Farrell and Ioannou, 1996a, b; Achatz and Schmitz, 2005a, b) indicates that initial conditions can easily determine the dominant finite-amplitude response for flows that are nonorthonormal (i.e., eigenvectors are not orthogonal, hence any one solution is not described by a unique combination), as is the case for all sheared and stratified flows. In simple terms, a flow that does not have orthogonal eigenfunctions can experience rapid growth of arbitrary initial perturbations that project onto eigenfunctions having similar structures but very different growth (or decay!) rates. Indeed, the underlying flow may be stable from the perspective of traditional linear stability analysis, but nevertheless allow transitions to instability and turbulence that can only be understood from the more general optimal perturbation perspective.

4.1. Nonlinear wave-wave interactions

Wave-wave interactions have been explored extensively in seeking to understand spectral energy transfers and the maintenance and apparent universality of the GW spectrum in the oceans and the atmosphere (Mied, 1976; McComas and



Fig. 7. Instability growth rates as a function of streamwise (α) and spanwise (β) wavenumber for GWs having a = 0.7 (left) and 1.1 (right) and intrinsic frequencies $\omega = N/4, N/3, N/2$, and N/1.4 (top to bottom) at Re = 1000 (after Lombard and Riley, 1996). Wavenumbers are normalized by the primary GW total wavenumber.

Bretherton, 1977; Yeh and Liu, 1981; Müller et al., 1986; Dunkerton, 1989). Generalizations to include GW–vortical mode interactions, the influences of

mean shear, and the links of these "resonant" threewave interactions at small amplitudes to local GW instabilities at larger amplitudes have also been explored more recently (Dong and Yeh, 1988; Yeh and Dong, 1989; Klostermeyer, 1991; Vanneste, 1995; Sonmor and Klaassen, 1997). It is important to recognize, however, that these and other nonlinear dynamics do not occur in isolation, but as part of a continuous competition. As an example, Müller et al. (1986) examined the relevance of offresonant or higher-order interactions in describing GW spectral energy transfers. In another, Klostermeyer (1991) identified specific resonant interactions or instability processes across the full range of GW amplitudes, suggested that resonance dynamics may be the basis of all GW instability, and found with numerical studies that multiple interactions quickly populated the spectrum across a wide range of frequencies and wavenumbers. Finally, Thorpe (1994) found a parametric subharmonic instability (PSI) to operate effectively even at "unstable" GW amplitudes when viewed from a traditional instability perspective. Thus, wave-wave interaction dynamics must play a central role in GW propagation, interaction, and instability dynamics, though their full impacts remain to be assessed.

As an example of the interactions among and competition between instability processes, we show in Fig. 8 results of a direct numerical simulation (DNS) of the evolution of a GW having an amplitude of a = u'/(c - U) = 0.7, well below that required for convective instability in the traditional view of GW instability (though anticipated to be unstable to wave-wave interactions in the analyses by Lombard and Riley (1996) and Sonmor and Klaassen (1997)). In this simulation, we have inclined the simulation domain along the phase of the GW, which is propagating upward and leftward (the long edge of the domain is parallel to the phase surfaces of the GW).

The images on the left are streamwise cross sections in the plane of propagation; those on the right are spanwise cross sections (a plane perpendicular to the group velocity) at the center of the slanted domain. In both set of images, bright values denote high shear or vorticity and black shades are zero shear or vorticity, and times are in units of T_b since initial conditions were posed. In the upper images, only the primary GW is apparent, as the initial noise is very small. In the second images, there is no detectable spanwise (right image) variation, but there are now significant variations in the streamwise structure (left image) that indicate that (2D) wave–wave interactions have begun to play a role in the evolution. The apparent orienta-



Fig. 8. A DNS of GW instability for a = 0.7, Re = 3000, and $\omega = N/3$ in a computational domain slanted along the GW phase surfaces. Time units are buoyancy periods, $T_{\rm b}$, The dominant instability initially is a largely 2D wave–wave interaction, but increasing gradients arising from wave superposition trigger a local 3D instability thereafter. At later times, the flow is returning to a quasi-2D structure, with largely horizontal phase surfaces.

tion of the emerging wave structures are more nearly horizontal than the initial GW and are consistent with a PSI as initially identified by McComas and Bretherton (1977) and observed thereafter in numerical and laboratory studies by Klostermeyer (1991), Thorpe (1994), Vanneste (1995), and others. The subharmonics that arise (typically having larger m and smaller $|\mathbf{k}|$ than the initial GW) must have periodic boundary conditions in the computational domain (not multiple horizontal wavelengths, but subharmonics, i.e. λ/n , with n an integer), but appear not strongly constrained by this requirement. In studies currently in progress, we are relaxing this constraint to determine whether it has restricted the form and growth rate of the instability.

The evolution departs sharply from a largely 2D flow by the third images, where we now see that the superposition of 2D motions having smaller vertical scales has led to larger gradients and local 3D instability as observed in our earlier GW breaking simulations of a single wave (Andreassen et al., 1994, 1998; Fritts et al., 1994, 1996b, 1998, 2003). Indeed, the local instability structure appears to have much in common with those earlier simulations, especially the streamwise-aligned and counter-rotating nature of the initial instability (some of this character can still be seen in the spanwise cross section at t = 66).

By the fourth images, the instability has largely disappeared, the GW amplitudes have been reduced, and the flow has returned to a more 2D character having largely subharmonic structure. The message from this simulation, like the laboratory study by Thorpe (1994), is that different instability dynamics can occur simultaneously, in competition with one another, or successively, depending on the structure of the large- and smallscale flows.

4.2. GW instability dynamics

Wave-wave interactions were seen above to lead to local instability as the 2D flow evolved larger gradients at smaller spatial scales. Here we show more quantitative results for a single monochromatic GW in order to understand more fully the implications of wave breaking for GW amplitude evolution. A similar result was discussed in some detail by Fritts et al. (2003) previously. However, the present results are for a significantly higher Reynolds number and exhibit more vigorous turbulence extending to smaller scales of motion.

Streamwise and spanwise cross sections of vorticity for a simulation of GW breaking for an initial amplitude $a = 1.1, \omega = N/3$, and Re = 3000 are shown in Fig. 9. Time units, the initial noise spectrum, and the domain size are the same as in Fig. 8. The first thing to note is that instability both arises and advances much more rapidly for a = 1.1than for a = 0.7. The first images for a = 0.7 in Fig. 8 show no evidence of instability, whereas the upper right (spanwise) image in Fig. 9 exhibits significant modulation within the upper region of high vorticity at about the same time. The second images shown less than a wave period later exhibit well-developed instability features having displacements extending across a significant part of the GW phase structure. The form of the instability is the same as identified by Fritts et al. (2003) at a Reynolds number, Re = 1000, comprising streamwise-aligned counter-rotating vortices that are largely horizontal, largely confined to the most unstable phase of the GW, and linked by loops



Fig. 9. As in Fig. 8, but for an initial GW amplitude of a = 1.1. In this case, initial instability is 3D, with counter-rotating, streamwise-aligned vortices dominating the early evolution, and vigorous turbulence arises quickly (within $\sim a$ wave period of strong initial instability). The late wave field includes the initial GW at smaller amplitude and other components having more nearly horizontal phase alignment.

having largely spanwise vorticity and extending lower in the GW structure. Another $2.5T_b$ later (at t = 12.7), the large-scale coherent structures have been replaced by intense turbulence seeming to fill the majority of the GW field and to include a train of coherent spanwise vortices along one phase of the GW. Another $\sim 10T_b$ later, the turbulence has largely abated, and there is evidence of a residual initial GW as well as more nearly horizontal structures similar to those seen for a = 0.7 in Fig. 8.

The GW amplitude and heat flux throughout this evolution are shown in the left panel of Fig. 10. The amplitude is seen to drop from its initial value by more than a factor of 3 to $a\sim0.35$, with most of this reduction occurring between the second and third images in Fig. 9. Thereafter, the amplitude (measured in the velocity field) oscillates with half the GW period due to continuing exchanges of GW energy between kinetic and potential. The heat flux computed throughout the simulation is seen to be zero prior to wave breaking, to peak sharply during breaking, and to oscillate about zero again, also at



Fig. 10. Horizontal velocity amplitude, u' (left panel, solid line, left axis), and heat flux, w'T, averaged over the GW phase (left panel, dashed line, right axis) throughout GW breaking for an initial amplitude of a = 1.1. Note that breaking and amplitude reduction occur within one GW period. Oscillations at later stages exhibit exchanges between kinetic and potential energy at a period half that of the primary GW. The right panel shows the primary GW amplitude evolution for various initial amplitudes all having Re = 1000 and $\omega = N/3$.

half the GW period, following GW amplitude reduction and accompanying turbulence decay.

The temporal evolutions of the GW amplitudes for the two simulations discussed above, as well as for a = 0.9 and 0.5, are shown together in the right panel of Fig. 10. At first glance, these results are startling. The larger initial GW amplitudes (a = 1.1)and 0.9) exhibit earlier and faster decay to final amplitudes of $a \sim 0.35$ or less. We earlier expressed surprise that the GW amplitude would decay so sharply relative to expectations of simple linear theory (Fritts et al., 2003). But is also surprising that an initial amplitude of a = 0.9 decays to an equivalent final amplitude. Even more surprising is the amplitude decay observed at smaller initial amplitudes. Indeed, these decay to zero on very long time scales. The reason is that in these cases, the decay process is wave-wave interaction rather than breaking and turbulence dissipation, and wavewave interactions apparently transfer all of the initial GW energy to other components of the motion field on long time scales. Note, however, that while the time scale for instability at larger initial amplitudes is short and suggests that these dynamics will be important in the middle atmosphere, the time scale for wave-wave interaction is much longer. Hence these dynamics will likely play a role where vertical scales and group velocities are small, or in competition with local instability dynamics at larger amplitudes, as noted by Thorpe (1994). But they are unlikely to compete at larger group velocities with other dynamics accompanying rapid amplitude growth due to density decay with increasing altitude.



Fig. 11. GW amplitude evolutions for a = 0.7, Re = 1000, and $\omega = N/3$ for differing computational domain depths. Note that larger domains enable different and earlier instabilities, but that the apparent instability evolution and ultimate GW amplitude is a function of the allowed spatial scales (and resonant interactions).

A final caveat is illustrated in Fig. 11. Displayed here are the amplitude evolutions for three simulations having an initial amplitude a = 0.7, and all other parameters identical except for the depth of the computational domain. To explore sensitivity to the presence of other allowed scales (and increasing degrees of freedom), we first doubled, then tripled, the domain depth. In the first case, we observed an earlier and more rapid amplitude decay than in the "standard" domain, also to a final zero amplitude, but now having amplitude oscillations of period $\sim 7T_{\rm b}$, suggesting an oscillation between kinetic and potential energy of a wave motion having a period $\sim 14T_{\rm b}$. This itself is not surprising, as increasing the domain size (and degrees of freedom) allows for other possible interactions than can occur in the more confined domain. The surprising result is that when the domain depth is tripled, we obtain a dramatically different solution than for either of the smaller domains. Again, instability is advanced to an earlier time, relative to the standard domain (but not to the same degree), and initial oscillations occur suggesting the excitation of a GW having a period of $\sim 14T_{\rm b}$, as in the doubled domain. In this case, however, we also see that the initial GW amplitude no longer falls to zero, but to an intermediate amplitude of $a \sim 0.22$, and that oscillations occur at ~half the initial GW period as the amplitude stabilizes at the smaller value. This suggests that the initial wave-wave interaction was interrupted, that local instability occurred, and that the resulting drop in amplitude prevented further energy transfer by the initial wave-wave interaction. While we are working to understand these results more fully, they suggest that care must be taken in assessing instability dynamics and effects in general, and that one must be aware of possible artificial constraints on available modes of instability.

4.3. Wave-mean flow interactions

Body forcing accompanying GW dissipation and momentum flux divergence is now understood to be responsible for closure of the mesospheric jets and reversal of the meridional temperature gradient at the mesopause, to play a role in the equatorial quasi-biennial (QBO) and semiannual oscillations (SAO), and to contribute to systematic departures from geostrophic or gradient wind balance under a variety of other conditions based on over two decades of modeling, theoretical, and observational studies. At middle and high latitudes, GW momentum flux divergence contributes a mean body forcing of \sim 50–100 m s⁻¹ day⁻¹ near the mesopause that opposes the summer and winter mesospheric jets, alters the mean force balance, and results in a mean meridional motion of $\sim 20 \,\mathrm{m \, s^{-1}}$ that yields a Coriolis torque that balances the zonal body force (Nastrom et al., 1982; Holton, 1982, 1983; Garcia, 1989; Fritts and Luo, 1995). Observations and theory also attribute the majority of this forcing at middle and high latitudes to GW having relatively high intrinsic frequencies due to the shape of the frequency spectra displayed in Fig. 5 (Vincent and Reid, 1983; Fritts, 1984a; Fritts and Vincent, 1987).

The latitudinal, seasonal, and hemispheric variations of the mean GW forcing are not known well, but observations of mean meridional motions and the mean solstice thermal structure, and especially the separated mesopause, with a low, cold mesopause confined to middle and high latitudes of the summer hemisphere, offer important insights (Nastrom et al., 1982; Wang and Fritts, 1990; von Zahn et al., 1996; Lübken, 1999). Additional insights on inter-hemispheric variability in GW forcing are provided by emerging, but still controversial, lidar, radar, rocket, and satellite measurements of temperatures, winds, polar mesospheric clouds (PMC), and polar mesosphere summer echoes (PMSE) (Balsley et al., 1993, 1995; Lübken et al., 1999; Gardner et al., 2001; Bailey et al., 2005). Similar body forcing, though much smaller in magnitude, also occurs at lower altitudes, in particular the winter polar stratosphere (Hitchman et al., 1989; Garcia and Boville, 1994), and at lower latitudes, where GW momentum flux divergence contributes to the structure and variability of the QBO and SAO and is associated largely with GWs excited by deep convection across a wide range of frequencies (Dunkerton, 1982, 1997; Salby and Garcia, 1987; Bergman and Salby, 1994; Mayr et al., 1997; Baldwin et al., 2001; Pfister et al., 1993a, b; Alexander et al., 2000; Alexander and Vincent, 2000).

A useful framework from which to view the mean zonal forcing is via the "downward control principle", which relates the Lagrangian vertical motion to the meridional gradient of the zonal body force applied to the atmosphere at higher altitudes (see McIntyre, 1989; Haynes et al., 1991; Garcia and Boville, 1994). Essentially, this describes the circulation through any level that must occur to satisfy continuity and accommodate the vertical coupling due to GWs arising in the lower atmosphere and dissipating at higher altitudes. Hence, a mean momentum flux at any altitude implies a corresponding mean vertical motion (the momentum flux must be zero at the pole) because a momentum flux divergence above requires a vertical mass flux to balance that accompanying the residual mean meridional motion providing the force balance at the altitude of GW dissipation.

Finally, recent efforts have accounted for the "missing forces" accompanying turning, or a change in GW propagation direction, in an environment having large-scale vertical vorticity. It is clear how to account for momentum transport and mean flow responses in cases where GW propagation remains confined to a vertical plane. It was not obvious, however, how to account for momentum transport in cases where GW propagation direction rotates in the horizontal. This problem was addressed by Bühler and McIntyre (2003, 2005), with the result that the momentum balance is just as we would have hoped: net momentum is conserved in two horizontal dimensions.

5. Summary and conclusions

A variety of measurements, modeling, and theoretical studies have provided an increasingly quantitative understanding of the mean forcing of the middle atmosphere by GWs in the last two decades. Only more recently, however, have we begun to appreciate the considerable variability of GWs and their effects at higher altitudes. This variability arises due to the inherent intermittency in GW sources, energy and momentum transports, propagation conditions, wave–wave and wave– mean flow interactions, and especially instability dynamics across the full range of GW spatial and temporal scales.

Mean forcing by GWs closes the mesospheric jets, induces a strong mean meridional motion, reverses the meridional temperature gradient in the upper mesosphere relative to radiative inputs, and results in a mesopause having two fairly distinct altitudes under solstice conditions. Similar, though smaller, responses also occur in the meridional (and vertical) circulation and thermal structure of the polar winter stratosphere. At lower latitudes, GWs contribute significantly to the mean structures of the QBO and SAO, with effects extending throughout the middle atmosphere.

GW variability also has major effects across a wide range of scales. At larger scales, these include modulation by and feedbacks on tidal and planetary wave motions and seasonal, inter-annual, and interhemispheric variations in mean zonal wind structure, the induced residual (meridional and vertical) circulation, and the mean temperature structure. At smaller scales, GWs contribute highly variable local structure, variances, local energy and momentum fluxes, additional GW excitation, and turbulence and diffusion. Indeed, we expect that as our understanding of GW dynamics improves further, we will find evidence of variability (and mean responses) extending to even higher altitudes that recognized at present.

Acknowledgments

Support for this research was provided by AFOSR under contract F49620-00-C-0045, NASA under contracts NAS5-02036 and NAS5-02069, and NSF under grants ATM-0137354, ATM-0307910, and ATM-0314060.

References

- Achatz, U., Schmitz, G., 2005a. Shear and convective instability of inertia-gravity wave packets: Modal and nonmodal growth. Journal of Atmospheric Science, submitted for publication.
- Achatz, U., Schmitz, G., 2005b. Optimal growth in convectively stable inertia-gravity wave packets: Three-dimensional structure and early nonlinear development. Journal of Atmospheric Science, submitted for publication.
- Alexander, M.J., Holton, J.R., Durran, D.R., 1995. The gravity wave response above deep convection in a squall line simulation. Journal of Atmospheric Science 52, 2212–2226.
- Alexander, M.J., Beres, J.H., Pfister, L., 2000. Tropical stratospheric gravity wave activity and relationship to clouds. Journal of Geophysical Research 105, 22,299–22,309.
- Alexander, M.J., Tsuda, T., Vincent, R.A., 2002. Latitudinal variations observed in gravity waves with short vertical wavelengths. Journal of Atmospheric Science 59, 1394–1404.
- Alexander, M.J., Vincent, R.A., 2000. Gravity waves in the tropical lower stratosphere: a model study of seasonal and interannual variability. Journal of Geophysical Research 105, 22299–22310.
- Allen, S.J., Vincent, R.A., 1995. Gravity wave activity in the lower atmosphere: seasonal and latitudinal variations. Journal of Geophysical Research 100, 1327–1350.
- Andreassen, O., Wasberg, C.-E., Fritts, D.C., Isler, J.R., 1994. Gravity wave breaking in two and three dimensions, 1. Model description and comparison of two-dimensional evolutions. Journal of Geophysical Research 99, 8095–8108.
- Andreassen, O., Hvidsten, P.O., Fritts, D.C., Arendt, S., 1998. Vorticity dynamics in a breaking gravity wave, 1. Initial instability evolution. Journal of Fluid Mechanics 367, 27–46.
- Andrews, D.G., McIntyre, M.E., 1976. Planetary waves in horizontal and vertical shear: the generalized Eliassen–Palm relation and the mean zonal acceleration. Journal of Atmospheric Science 33, 2031–2048.
- Bailey, S.M., Merkel, A.W., Thomas, G.E., 2005. Observations of polar mesospheric clouds by the Student Nitric Oxide Explorer. Journal of Geophysical Research 109, D13203, doi:10.1029/2004JD005422.
- Baldwin, M.P., Gray, L.J., Dunkerton, T.J., Hamilton, K., Haynes, P.H., Randel, W.J., Holton, J.R., Alexander, M.J., Hirota, I., Horinouchi, T., Jones, D.B.A., Kinnersley, J.S., Marquardt, C., Sato, K., Takahashi, M., 2001. The quasibiennial oscillation. Reviews of Geophysics 39, 179–229.
- Balsley, B.B., Woodman, R.F., Sarango, M., Urbino, J., Rodriguez, R., Ragaini, E., Carey, J., 1993. Southern hemisphere PMES: where are they? Geophysical Research Letters 20, 1983–1985.

- Balsley, B.B., Woodman, R.F., Sarango, M., Rodriguez, R., Urbino, J., Ragaini, E., Carey, J., Huaman, J., Giraldez, A., 1995. On the lack of southern-hemisphere PMSE. Journal of Geophysical Research 100, 11685–11693.
- Bergman, J.W., Salby, M.L., 1994. Equatorial wave activity derived from fluctuations in observed convection. Journal of Atmospheric Science 51, 3791–3806.
- Broutman, D., 1986. On internal wave caustics. Journal of Physical Oceanography 16, 1625–1635.
- Broutman, D., Young, W.R., 1986. On the interaction of smallscale internal waves with near-inertial waves. Journal of Fluid Mechanics 166, 341–358.
- Broutman, D., MaCaskill, C., McIntyre, M.E., Rottman, J.W., 1997. On Doppler spreading models of internal waves. Geophysical Research Letters 24, 2813–2816.
- Broutman, D., Rottman, J.W., Eckermann, S.D., 2002. Maslov's method for stationary hydrostatic mountain waves. Quarterly Journal of Royal Meteorological Society 128, 1159–1172.
- Broutman, D., Rottman, J.W., Eckermann, S.D., 2003. A simplified Fourier method for nonhydrostatic mountain waves. Journal of Atmospheric Science 60, 2686–2696.
- Broutman, D., Rottman, J.W., Eckermann, S.D., 2004. Ray methods for internal waves in the atmosphere and ocean. Annual Review on Fluid Mechanics 36, 233–253.
- Bruhwiler, D.L., Kaper, T.J., 1995. Wavenumber transport: scattering of small-scale internal waves by large-scale wavepackets. Journal of Fluid Mechanics 289, 379–405.
- Buckley, G., Broutman, D., Rottman, J.W., Eckermann, S., 1999. On the importance of weak steady shear in the refraction of short internal waves. Geophysical Research Letters 26, 2877–2880.
- Bühler, O., McIntyre, M.E., 1999. On shear generated gravity waves that reach the mesosphere, part II: Wave propagation. Journal of Atmospheric Science 56, 3764–3773.
- Bühler, O., McIntyre, M.E., 2003. Remote recoil: a new wave-mean interaction effect. Journal of Fluid Mechanics 492, 207–230.
- Bühler, O., McIntyre, M.E., 2005. Wave capture and wave-vortex duality. Journal of Fluid Mechanics 534, 67–95.
- Bühler, O., McIntyre, M.E., Scinocca, J.F., 1999. On shear generated gravity waves that reach the mesosphere, part I: Wave generation. Journal of Atmospheric Science 56, 3749–3763.
- Chimonas, G., Grant, J.R., 1984. Shear excitation of gravity waves. Part II: upscale scattering from Kelvin–Helmholtz waves. Journal of Atmospheric Science 41, 2278–2288.
- Clark, T.L., Hauf, T., Kuettner, J.P., 1986. Convectively forced internal gravity waves: results from two-dimensional numerical experiments. Quarterly Journal of Royal Meteorological Society 112, 899–925.
- Dewan, E.M., Picard, R.H., O'Neil, R.R., Gardiner, H.A., Gibson, J., Mill, J.D., Richards, E., Kendra, M., Gallery, W.O., 1998. MSX satellite observations of thunderstormgenerated gravity waves in mid-wave infrared images of the upper Stratosphere. Geophysical Research Letters 25, 939–942.
- Dong, B., Yeh, K.C., 1988. Resonant and nonresonant wavewave interactions in an isothermal atmosphere. Journal of Geophysical Research 93, 3729–3744.
- Dunkerton, T.J., 1982. Theory of the mesopause semiannual oscillation. Journal of Atmospheric Science 39, 2681–2690.

- Dunkerton, T.J., 1987. Effect of nonlinear instability on gravity wave momentum transport. Journal of Atmospheric Science 44, 3188–3209.
- Dunkerton, T.J., 1989. Theory of internal gravity wave saturation. Pure and Applied Geophysics 130, 373–397.
- Dunkerton, T.J., 1997. The role of gravity waves in the quasibiennial oscillation. Journal of Geophysical Research 102, 26053–26076.
- Eckermann, S.D., 1997. Influence of wave propagation on the Doppler-spreading of atmospheric gravity waves. Journal of Atmospheric Science 54, 2554–2573.
- Farrell, B.F., Ioannou, P.J., 1996a. Generalized stability theory. Part I: autonomous operators. Journal of Atmospheric Science 53, 2025–2040.
- Farrell, B.F., Ioannou, P.J., 1996b. Generalized stability theory. Part II: nonautonomous operators. Journal of Atmospheric Science 53, 2041–2053.
- Fetzer, E.J., Gille, J.C., 1994. Gravity wave variance in LIMS temperatures. Part I. Variability and comparison with background winds. Journal of Atmospheric Science 51, 2461–2483.
- Fetzer, E.J., Gille, J.C., 1996. Gravity wave variance in LIMS temperatures. Part II: Comparison with the zonal-mean momentum balance. Journal of Atmospheric Science 53, 398–410.
- Fovell, R., Durran, D., Holton, J.R., 1992. Numerical simulations of convectively generated stratospheric gravity waves. Journal of Atmospheric Science 49, 1427–1442.
- Fritts, D.C., 1984a. Gravity wave saturation in the middle atmosphere: a review of theory and observations. Review of Geophysics and Space Physics 22, 275–308.
- Fritts, D.C., 1984b. Shear excitation of atmospheric gravity waves. Part II: nonlinear radiation from a free shear layer. Journal of Atmospheric Science 41, 524–537.
- Fritts, D.C., 1989. A review of gravity wave saturation processes, effects, and variability in the middle atmosphere. Pure and Applied Geophysics 130, 343–371.
- Fritts, D.C., Alexander, M.J., 2003. Gravity dynamics and effects in the middle atmosphere. Reviews of Geophysics 41.
- Fritts, D.C., Luo, Z., 1992. Gravity wave excitation by geostrophic adjustment of the jet stream, Part 1: twodimensional forcing. Journal of Atmospheric Science 49, 681–697.
- Fritts, D.C., Luo, Z., 1995. Dynamical and radiative forcing of the summer mesopause circulation and thermal structure, 1. Mean Solstice conditions. Journal of Geophysical Research 100, 3119–3128.
- Fritts, D.C., Nastrom, G.D., 1992. Sources of mesoscale variability of gravity waves, II: frontal, convective, and jet stream excitation. Journal of Atmospheric Science 49, 111–127.
- Fritts, D.C., VanZandt, T.E., 1987. Effects of Doppler shifting on the frequency spectra of atmospheric gravity waves. Journal of Geophysical Research 92, 9723–9732.
- Fritts, D.C., VanZandt, T.E., 1993. Spectral estimates of gravity wave energy and momentum fluxes, I: energy dissipation, acceleration, and constraints. Journal of Atmospheric Science 50, 3685–3694.
- Fritts, D.C., Vincent, R.A., 1987. Mesospheric momentum flux studies at Adelaide, Australia: observations and a gravity wave/tidal interaction model. Journal of Atmospheric Science 44, 605–619.

- Fritts, D.C., Isler, J.R., Andreassen, O., 1994. Gravity wave breaking in two and three dimensions, 2. Three-dimensional evolution and instability structure. Journal of Geophysical Research 99, 8109–8123.
- Fritts, D.C., Palmer, T.L., Andreassen, O., Lie, I., 1996a. Evolution and breakdown of Kelvin–Helmholtz billows in stratified compressible flows, I: comparison of two- and threedimensional flows. Journal of Atmospheric Science 53, 3173–3191.
- Fritts, D.C., Garten, J.F., Andreassen, O., 1996b. Wave breaking and transition to turbulence in stratified shear flows. Journal of Atmospheric Science 53, 1057–1085.
- Fritts, D.C., Arendt, S., Andreassen, O., 1998. Vorticity dynamics in a breaking internal gravity wave, 2. Vortex interactions and transition to turbulence. Journal of Fluid Mechanics 367, 47–65.
- Fritts, D.C., Vadas, S.A., Yamada, Y., 2002. An estimate of strong local gravity wave body forcing based on OH airglow and meteor radar observations. Geophysical Research Letters 29 (10), 1.
- Fritts, D.C., Bizon, C., Werne, J.A., Meyer, C.K., 2003. Layering accompanying turbulence generation due to shear instability and gravity wave breaking. Journal of Geophysical Research 108 (D8), 8452.
- Garcia, R.R., 1989. Dynamics, radiation, and photochemistry in the mesosphere: implications for the formation of noctilucent clouds. Journal of Geophysical Research 94, 14,605–14,615.
- Garcia, R.R., Boville, B., 1994. "Downward control" of the mean meridional circulation and temperature distribution of the polar winter stratosphere. Journal of Atmospheric Science 51, 2238–2245.
- Garcia, R.R., Sassi, F., 1999. Modulation of the Mesospheric Semiannual Oscillation by the Quasibiennial Oscillation. Earth Planets Space 51, 563–569.
- Gardner, C.S., Papen, G.C., Chu, X., Pan, W., 2001. First lidar observations of middle atmosphere temperatures, Fe densities, and polar mesospheric clouds over the North and South Poles. Geophysical Research Letters 28, 1199–1202.
- Griffiths, M., Reeder, M.J., 1996. Stratospheric inertia-gravity waves generated in a numerical model of frontogenesis. I: model solutions. Quarterly Journal of Royal Meteorological Society 122, 1153–1174.
- Haynes, P.H., Marks, C.J., McIntyre, M.E., Shephard, T.G., Shine, K.P., 1991. On the "Downward Control" of extratropical diabatic circulations by eddy-induced mean zonal forces. Journal of Atmospheric Science 48, 657–678.
- Hecht, J.H., Liu, A.Z., Walterscheid, R.L., Rudy, R.J., 2005. Maui Mesosphere and Lower Thermosphere (Maui MALT) observations of the evolution of Kelvin–Helmholtz billows formed near 86 km altitude. Journal of Geophysical Research 109, D22S10, doi:10.1029/2003JD003908.
- Hertzog, A., Vial, F., 2001. A study of the dynamics of the equatorial lower atmosphere by use of ultra-long duration balloons, 2, gravity waves. Journal of Geophysical Research 106, 22,745–22,761.
- Hines, C.O., 1971. Generalizations of the Richardson criterion for the onset of atmospheric turbulence. Quarterly Journal of Royal Meteorological Society 97, 429–439.
- Hines, C.O., 1988. The generation of turbulence by atmospheric gravity waves. Journal of Atmospheric Science 45, 1269–1278.
- Hitchman, M.H., Gille, J.C., Rodgers, C.D., Brasseur, G., 1989. The separated polar winter stratopause: a gravity wave driven

climatological feature. Journal of Atmospheric Science 46, 410-422.

- Holton, J.R., 1982. The role of gravity wave induced drag and diffusion in the momentum budget of the mesosphere. Journal of Atmospheric Science 39, 791–799.
- Holton, J.R., 1983. The influence of gravity wave breaking on the general circulation of the middle atmosphere. Journal of Atmospheric Science 40, 2497–2507.
- Horinouchi, T., Nakamura, T., Kosaka, J., 2002. Convectively generated mesoscale gravity waves simulated throughout the middle atmosphere. Geophysical Research Letters 21, 1.
- Isler, J.R., Fritts, D.C., 1996. Gravity wave variability and interaction with lower-frequency motions in the mesosphere and lower thermosphere over Hawaii. Journal of Atmospheric Science 53, 37–48.
- Jiang, J.H., Wu, D.L., Eckermann, S.D., 2002. Upper Atmosphere Research Satellite (UARS) MLS observation of mountain waves over the Andes. Journal of Geophysical Research 107 (D20), 1.
- Jiang, J.H., Wu, D.L., Eckermann, S.D., Ma, J., 2003. Mountain waves in the middle atmosphere: microwave limb sounder observations and analyses. Advances in Space Research 32 (5), 801–806.
- Jiang, J.H., Eckermann, S.D., Wu, D.L., Ma, J., 2004. A search for mountain waves in MLS stratospheric limb radiance from the Northern Hemisphere: data analysis and global mountain wave modeling. Journal of Geophysical Research 109, D03107.
- Karoly, D.J., Roff, G.L., Reeder, M.J., 1996. Gravity wave activity associated with tropical convection detected in TOGA COARE sounding data. Geophysical Research Letters 23, 261–264.
- Kim, Y.-J., Eckermann, S.D., Chun, H.-Y., 2003. A overview of the past, present and future of gravity-wave drag parameterization for numerical climate and weather prediction models. Atmosphere-Ocean 41, 65–98.
- Klostermeyer, J., 1991. Two- and three-dimensional parametric instabilities in finite amplitude internal gravity waves. Geophysical and Astrophysical Fluid Dynamics 64, 1–25.
- Lane, T.P., Clark, T.L., 2002. Gravity waves generated by the dry convective boundary layer: two-dimensional scale selection and boundary layer feedback. Quarterly Journal of Royal Meteorological Society 128, 1543–1570.
- Lane, T.P., Reeder, M.J., Clark, T.L., 2001. Numerical modeling of gravity wave generation by deep tropical convection. Journal of Atmospheric Science 58, 1249–1274.
- Lighthill, J., 1978. Waves in Fluids. Cambridge University Press, Cambridge 504pp.
- Lombard, P.N., Riley, J.J., 1996. Instability and breakdown of internal gravity waves. I. Linear stability analysis. Physics of Fluids 8, 3271–3287.
- Lübken, F.-J., 1999. Thermal structure of the Arctic summer mesosphere. Journal of Geophysical Research 104, 9135–9149.
- Lübken, F.-J., Jarvis, M.J., Jones, G.O.L., 1999. First in situ temperature measurements at the Antarctic summer mesopause. Geophysical Research Letters 26, 3581–3584.
- Luo, Z., Fritts, D.C., 1993. Gravity wave excitation by geostrophic adjustment of the jet stream, Part II: three-dimensional forcing. Journal of Atmospheric Science 50, 104–115.
- Manson, A.H., Meek, C.E., Hall, G.E., 1998. Correlations of gravity waves and tides in the mesosphere over Saskatoon.

Journal of Atmospheric and Solar-Terrestrial Physics 60, 1089–1107.

- Mayr, H.G., Mengel, J.G., Hines, C.O., Chan, K.L., Arnold, N.F., Reddy, C.A., Porter, H.S., 1997. The gravity wave Doppler spread theory applied in a numerical spectral model of the middle atmosphere 2. Equatorial oscillations. Journal of Geophysical Research 102, 26,093–26,105.
- McComas, C.H., Bretherton, F.P., 1977. Resonant interaction of oceanic internal waves. Journal of Geophysical Research 82, 1397–1412.
- McIntyre, M.E., 1989. On dynamics and transport near the polar mesopause in summer. Journal of Geophysical Research 94, 14,617–14,628.
- McIntyre, M.E., 2003. In: Holton, J.R., Pyle, J.A., Curry, J.A. (Eds.), Balanced Flow, Encyclopedia of Atmospheric Sciences, vol. 2. Academic Press, New York, pp. 680–685.
- Mied, R.P., 1976. The occurrence of parametric instabilities in finite amplitude internal gravity waves. Journal of Fluid Mechanics 78, 763–784.
- Müller, P., Holloway, G., Henyey, F., Pomphrey, N., 1986. Nonlinear interactions among internal gravity waves. Reviews of Geophysics 24, 493–536.
- Nakamura, T., Fritts, D.C., Isler, J.R., Tsuda, T., Vincent, R.A., 1997. Short-period fluctuations of the diurnal tide observed with low-latitude MF and meteor radars during CADRE: evidence for gravity wave/tidal interactions. Journal of Geophysical Research 102, 26,225–26,238.
- Nastrom, G.D., Balsley, B.B., Carter, D.A., 1982. Mean meridional winds in the mid- and high-latitude summer mesosphere. Geophysical Research Letters 9, 139–142.
- Nastrom, G.D., VanZandt, T.E., Warnock, J.M., 1997. Vertical wavenumber spectra of wind and temperature from highresolution balloon soundings in the lower atmosphere over Illinois. Journal of Geophysical Research 102, 6685–6702.
- Palmer, T.L., Fritts, D.C., Andreassen, O., 1996. Evolution and breakdown of Kelvin–Helmholtz billows in stratified compressible flows, II: instability structure, evolution, and energetics. Journal of Atmospheric Science 53, 3192–3212.
- Pfister, L., Starr, W., Craig, R., Loewenstein, M., 1986. Smallscale motions observed by aircraft in the tropical lower stratosphere: evidence for mixing and its relationship to largescale flows. Journal of Atmospheric Science 43, 3210–3225.
- Pfister, L., Scott, S., Loewenstein, M., Bowen, S., Legg, M., 1993a. Mesoscale disturbances in the tropical stratosphere excited by convection: observations and effects on the stratospheric momentum budget. Journal of Atmospheric Science 50, 1058–1075.
- Pfister, L., Chan, K.R., Bui, T.P., Bowen, S., Legg, M., Gary, B., Kelly, K., Proffitt, M., Starr, W., 1993b. Gravity waves generated by a tropical cyclone during the STEP tropical field program: a case study. Journal of Geophysical Research 98, 8611–8638.
- Piani, C., Durran, D., Alexander, M.J., Holton, J.R., 2000. A numerical study of three-dimensional gravity waves triggered by deep tropical convection and their role in the dynamics of the QBO. Journal of Atmospheric Science 57, 3689–3702.
- Preusse, P., Dörnbrack, A., Eckermann, S.D., Riese, M., Schaeler, B., Bacmeister, J.T., Broutman, D., Grossmann, K.U., 2002. Space-based measurements of stratospheric mountain waves by CRISTA, 1, Sensitivity, analysis method, and a case study. Journal of Geophysical Research 107 (D23), 8178, doi:10.1029/2001JD000699.

- Reeder, M.J., Griffiths, M., 1996. Stratospheric inertia-gravity waves generated in a numerical model of frontogenesis. II: wave sources and generation mechanisms. Quarterly Journal of Royal Meteorological Society 122, 1175–1195.
- Salby, M.L., Garcia, R.R., 1987. Transient response to localized episodic heating in the tropics. Part I: excitation and shorttime near-field behaviour. Journal of Atmospheric Science 44, 458–498.
- Scinocca, J.F., Ford, R., 2000. The nonlinear forcing of largescale internal gravity waves by stratified shear instability. Journal of Atmospheric Science 57, 653–672.
- Sentman, D.D., Wescott, E.M., Picard, R.H., Winick, J.R., Stenbaek-Nielson, H.C., Dewan, E.M., Moudry, D.R., São Sabbas, F.T., Heavner, M.J., 2003. Simultaneous observation of mesospheric gravity waves and sprites generated by a Midwestern thunderstorm. Journal of Atmospheric and Solar-Terrestrial Physics 65, 537–550.
- Shimizu, A., Tsuda, T., 1997. Characteristics of Kelvin waves and gravity waves observed with radiosondes over Indonesia. Journal of Geophysical Research 102, 26159–26171.
- Smith, A.K., 1996. Longitudinal variations in mesospheric winds: evidence for gravity wave filtering by planetary waves. Journal of Atmospheric Science 53, 1156–1173.
- Song, I.-S., Chun, H.-Y., Lane, P.P., 2003. Generation mechanisms of convectively forced internal gravity waves and their propagation to the stratosphere. Journal of Atmospheric Science 60, 1960–1980.
- Sonmor, L.J., Klaassen, G.P., 1997. Toward a unified theory of gravity wave stability. Journal of Atmospheric Science 54, 2655–2680.
- Sonmor, L.J., Klaassen, G.P., 2000. Mechanisms of gravity wave focusing in the middle atmosphere. Journal of Atmospheric Science 57, 493–510.
- Sutherland, B.R., 1999. Propagation and reflection of internal waves. Physics of Fluids 11, 1081–1090.
- Sutherland, B.R., 2000. Internal wave reflection in uniform shear. Quarterly Journal of Royal Meteorological Society 126, 3255–3287.
- Sutherland, B.R., 2001. Finite-amplitude internal wavepacket dispersion and breaking. Journal of Fluid Mechanics 429, 343–380.
- Taylor, M.J., Hapgood, M.A., 1988. Identification of a thunderstorm as a source of short period gravity waves in the upper atmospheric nightglow emissions. Planetary and Space Science 36, 975.
- Thayaparan, T., Hocking, W.K., MacDougall, J., 1995. Observational evidence of tidal/gravity wave interactions using the UWO 2 MHz radar. Geophysical Research Letters 22, 373–376.
- Thorpe, S.A., 1994. Observations of parametric instability and breaking waves in an oscillating tilted tube. Journal of Fluid Mechanics 261, 33–45.
- Tsuda, T., Murayama, Y., Wiryosumarto, H., Harijono, S.W.B., Kato, S., 1994. Radiosonde observations of equatorial atmosphere dynamics over Indonesia, 2, characteristics of gravity waves. Journal of Geophysical Research 99, 10,507–10,516.
- Tsuda, T., VanZandt, T.E., Mizumoto, M., Kato, S., Fukao, S., 1991. Spectral analysis of temperature and Brunt-Vaisala frequency fluctuations observed by radiosondes. Journal of Geophysical Research 96, 17265–17278.
- Tsuda, T., Nishida, M., Rocken, C., 2000. A global morphology of gravity wave activity in the stratosphere revealed by the

GPS occultation data (GPS/MET). Journal of Geophysical Research 105, 7257–7274.

- Vadas, S.L., Fritts, D.C., 2001. Gravity wave radiation and mean responses to local body forces in the atmosphere. Journal of Atmospheric Science 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 force s. Geophysical Research Letters 29 (20).
- Vadas, S.L., Fritts, D.C., 2004. Thermospheric responses to gravity waves arising from mesoscale convective complexes. Journal of Atmospheric and Solar-Terrestrial Physics 66, 781–804.
- Vadas, S.L., Fritts, D.C., Alexander, M.J., 2003. Mechanism for the generation of secondary waves in wave breaking regions. Journal of Atmospheric Science 60, 194–214.
- Vanneste, J., 1995. The instability of internal gravity waves to localized disturbances. Annals of Geophysics 13, 196–210.
- VanZandt, T.E., 1982. A universal spectrum of buoyancy waves in the atmosphere. Geophysical Research Letters 9, 575–578.
- Vincent, R.A., Alexander, M.J., 2000. Gravity waves in the tropical lower stratosphere: An observational study of seasonal and interannual variability. Journal of Geophysical Research 105, 17971–17982.
- Vincent, R.A., Reid, I.M., 1983. HF Doppler measurements of mesospheric momentum fluxes. Journal of Atmospheric Science 40, 1321–1333.
- von Zahn, U., Höffner, J., Eska, V., Alpers, M., 1996. The mesopause altitude: only two distinctive levels worldwide? Geophysical Research Letters 23, 3231–3234.
- Wada, K., Nitta, T., Sato, K., 1999. Equatorial inertia-gravity waves in the lower stratosphere revealed by TOGA-COARE IOP data. Journal of Meteorological Society, Japan 77, 721–736.

- Walterscheid, R.L., 1981. Inertio-gravity wave induced accelerations of mean flow having an imposed periodic component: Implications for tidal observations in the meteor region. Journal of Geophysical Research 86, 9698–9706.
- Walterscheid, R.L., 2000. Propagation of small-scale gravity waves through large-scale internal wave fields: Eikonal effects at low-frequency approximation critical levels. Journal of Geophysical Research 105, 18027–18037.
- Wang, D.-Y., Fritts, D.C., 1990. Mesospheric momentum fluxes observed by the MST radar at Poker Flat, Alaska. Journal of Atmospheric Science 47, 1511–1521.
- Wang, D.-Y., Fritts, D.C., 1991. Evidence of gravity wave-tidal interaction observed near the summer mesopause at Poker Flat, Alaska. Journal of Atmospheric Science 48, 572–583.
- Werne, J., Fritts, D.C., 1999. Stratified shear turbulence: evolution and statistics. Geophysical Research Letters 26, 439–442.
- Yamada, Y., Fukunishi, H., Nakamura, T., Tsuda, T., 2001. Breakdown of small-scale quasi-stationary gravity wave and transition to turbulence observed in OH airglow. Geophysical Research Letters 28, 2153–2156.
- Yeh, K.C., Dong, B., 1989. The nonlinear interaction of a gravity wave with the vortical modes. Journal of Atmospheric and Terrestrial Physics 51, 45–50.
- Yeh, K.C., Liu, C.H., 1981. The instability of atmospheric gravity waves through wave–wave interactions. Journal of Geophysical Research 86, 9722–9728.
- Zhong, L., Sonmor, L.J., Manson, A.H., Meek, C.E., 1995. The influence of time-dependent wind on gravity-wave propagation in the middle atmosphere. Annales Geophysicae 13, 375–394.
- Zhu, X., Holton, J.R., 1987. Mean fields induced by local gravitywave forcing in the middle atmosphere. Journal of Atmospheric Science 44, 620–630.