Gravity wave coupling from below: A review

Robert A. Vincent

School of Chemistry and Physics, University of Adelaide, Adelaide, SA 5005, Australia
E-mail: robert.vincent@adelaide.edu.au

Gravity waves efficiently couple energy and momentum from source regions in the lower atmosphere to the middle and upper atmosphere. Wave dissipation results in a drag that drives meridional circulations that can profoundly affect the state of the middle atmosphere. New developments in various observing techniques, as well as a better appreciation of the limits imposed by observational selection, allow for a better understanding of the role of gravity waves in the atmosphere. Here we summarise recent advances in the measurement of wave parameters, especially those obtained by satellite and in situ techniques. Of particular importance is the development of techniques to measure momentum fluxes on a hemispheric or even global scale. Improved understanding of gravity wave coupling processes into the thermosphere-ionosphere has resulted from better formulations of dissipation effects. Allied with new rapid incoherent scatter measurements it may soon be possible to identify sources of small and medium scale ionospheric travelling disturbances that are not linked to auroral phenomena.

1 Introduction

Gravity waves, or more accurately buoyancy waves, are ubiquitous in the atmosphere. Wave periods range from minutes to hours and spatial scales extend up to thousands of km. They are important because they efficiently couple energy and momentum from their source regions in the lower atmosphere to the middle and upper atmosphere. Their presence in the lower atmosphere has been noted for over one hundred years, but is only relatively recently that their importance in the middle and upper atmosphere has been appreciated.

Wavelike perturbations in the ionospheric F-region are often referred to as travelling ionospheric disturbances (TIDs). Both early and more recent studies of TIDs recognized that they are often not generated by geomagnetic disturbances but have possible links to meteorological sources in the lower atmosphere (e.g. Munro, 1958; Waldock and Jones, 1987). It was the realization that TIDs were the manifestation of gravity waves that led in part to the seminal paper by Hines (1960) whose pioneering work also helped establish the importance of wave transfer of energy and momentum from the lower to the middle and upper atmosphere (Hines, 1974).

The last decade has seen new techniques brought to bear in the study of gravity waves and their effects in the middle atmosphere and thermosphere-ionosphere.
Satellite observations provide a global perspective and wave effects can be observed in somewhat unexpected ways. For example, Fig. 1 shows lidar signals transmitted from the *Calipso* satellite that are backscattered from polar stratospheric clouds during a pass over the Antarctic Peninsula. Gravity waves modulate the cloud density, as indicated by the dashed lines, with the tilt of the constant phase surfaces characteristic of upward propagating waves. These wave effects in the middle stratosphere are probably manifestations of mountain waves produced by the strong winter eastward flow over the Antarctic Peninsula.

Transfer of momentum by gravity waves is particularly important. The quantity $\rho_o u' w'$, where $\rho_o$ is atmospheric density and $u'$ and $w'$ are the horizontal and vertical perturbation amplitudes, is conserved in the absence of wave dissipation. Momentum deposition associated with wave breaking or dissipation exerts a body force on the atmosphere (Andrews *et al.*, 1987). The zonal-average zonal wind momentum...
At mid- to high-latitudes at the solstices the acceleration of the mean zonal wind $\bar{u}$ is small and there is a balance between the Coriolis term and flux divergence. At the solstices the wave drag drives a pole-to-pole circulation in the mesosphere that profoundly changes the thermal structure of this region. The upwelling over the summer pole (Fig. 2) and the associated adiabatic cooling produces temperatures that can be as low as 100 K at heights near 85 km. The cold temperatures lead to freezing of even the small quantities of water vapour present in the mesosphere and this is manifest in such phenomena as noctilucent clouds and polar mesospheric summer echoes observed with ground-based radars (e.g. Rapp and Lübken, 2004). The associated sinking of air over the winter pole leads to a warming whose effects can be felt as low as 20 km (Garcia and Boville, 1994).

In the tropics, where $f \sim 0$, the second term in Eq. (1) is negligible and so gravity waves can drive an acceleration of the zonal-mean zonal wind. Hence gravity waves probably play important roles in driving the quasi-biennial oscillation (QBO) in the stratosphere and semiannual oscillations near the stratopause and mesopause (Dunkerton, 1997; Baldwin et al., 2001).

Full understanding of the role of gravity waves in the middle and upper atmosphere requires improved measurements on a global scale of wave parameters, especially momentum fluxes. Ideally these measurements should be made as a function of wave phase speed, so that their effects can be fully incorporated in numerical climate
models and vertical coupling effects properly understood. Hamilton (1999) provides a background to early gravity wave observations and theoretical developments while Fritts and Alexander (2003) provide a recent and comprehensive review of gravity wave theory and effects. Here we review recent observations of gravity waves made using a variety of techniques as well as coupling effects.

2 Observations
2.1 Observational selection
A wide variety of techniques, ranging from ground-based to in-situ balloon-borne measurements and space-based measurements, are used to study gravity waves and their characteristics. However, no technique is perfect; it is not possible for one method to measure wave parameters across the whole wave spectrum (Alexander, 1998). For example, Alexander and Barnet (2007) discuss the response of various satellite viewing systems, as illustrated in Fig. 3. Satellite instruments that view along the limb, such as HIRDLS, have excellent vertical resolution, but relatively poor horizontal resolution. In contrast, a nadir-viewing instrument such as AIRS has excellent horizontal resolution compared with the vertical. Alexander et al. (2008b) made a detailed investigation of the application of the GPS radio occultation technique to gravity wave studies. Their analysis shows that best retrievals occur for waves with quasi-horizontal phase surfaces, i.e. for waves with quasi-inertial frequencies. They conclude that caution must be used in using occultation data to retrieve GW parame-

Fig. 3. Illustration of space-based observations of gravity waves. The background colours show temperature perturbations of modelled gravity waves, while the ovals illustrate the weighting function cross-sections for various satellite viewing modes. Those on the left are for downward viewing satellites, while the ovals on the right are for limb viewing geometries. (From figure 3, Alexander and Barnet (2007). See their paper for more details.)
Gravity wave dispersion is another selection factor that needs to be taken into account. Just because a gravity wave packet is in the viewing region of a particular instrument it does not necessarily mean that accurate measurements of wave amplitude and fluxes can be made. Short-period gravity waves tend to propagate vertically much faster than longer period waves. Consequently, a high frequency wave packet will propagate up through a region of observation more quickly than a longer period packet (Alexander and Barnet, 2007) so that observations of long-period waves are favoured.

2.2 Satellite observations

The advent of microsatellites in low earth orbit, such as GPS/MET, CHAMP and the COSMIC constellation, has allowed gravity wave activity to be studied globally using the GPS occultation technique (e.g. Tsuda et al., 2000; de la Torre et al., 2004, 2006; Ratnam et al., 2004; Baumgaertner and McDonald, 2007; Hei et al., 2008). As a low earth orbit satellite sets, the signals from an occulting GPS satellite are refracted by the intervening atmosphere and ionosphere. The resulting ray bending gives information on the vertical distributions of electron density, temperature and, low in the atmosphere, humidity. Gravity wave potential energy per unit mass information is derived from the temperature profiles

\[
E_p = \frac{1}{2} \frac{g^2}{N^2} \frac{T' T_o}{T_o^2},
\]

where \(N\) is the Brunt frequency, \(g\) is the acceleration due to gravity and \(T'\) and \(T_o\) are the temperature perturbation and mean temperature, respectively. As shown in Fig. 4, all the necessary information to compute \(E_p\) can be derived from the GPS

Fig. 4. GPS estimates of gravity wave perturbation amplitudes. (From Baumgaertner and McDonald, A gravity wave climatology for Antarctica compiled from Challenging Minisatellite Payload/Global Positioning System (CHAMP/GPS) radio occultations, J. Geophys. Res., 112, D05103, 2007. Copyright 2007 American Geophysical Union. Reproduced by permission of American Geophysical Union.)
measurements, although as noted by Alexander et al. (2008b), caution is required in deriving GW parameters from occultation data.

Using several years of CHAMP data, Baumgaertner and McDonald (2007) were able to study the seasonal cycle in GW $E_p$ in the Antarctic lower stratosphere at heights between $\sim 10$ and 35 km. There is an annual cycle in $E_p$ with largest amplitudes in winter, a result similar to that reported by Zink and Vincent (2001) for the SH mid-latitude site of Macquarie Island. Baumgaertner and McDonald (2007) also reported an interesting longitudinal variation of wave activity, as well as a strong enhancement of wave energy at the edge of the polar vortex. Their results show a strong relationship between gravity wave activity and geographic location, indicating that topography is a strong source for wave activity, especially over the Antarctic Peninsula and the Southern Andes. Mountain waves from this region have even been linked to disturbances in the lower ionosphere (Hocke et al., 2002). On the other hand, an analysis by Hei et al. (2008) of 5 years of GPS data from the CHAMP satellite for both polar regions indicates the limited geographical extent of topographically forced waves and the importance of waves that they ascribe to gravity generation by planetary wave transience and/or breaking.

In the equatorial stratosphere, Tsuda et al. (2000), Alexander et al. (2002) and de la Torre et al. (2006) all report a maximum in GW $E_p$ extending between about 1°S and 10°N. However, the $E_p$ variations show significant geographic as well as interannual variability. de la Torre et al. (2006) found interannual enhancements in wave activity related to the QBO, similar to the findings of Vincent and Alexander (2000) made using radiosonde observations at Christmas Island (12°S) in the Indian Ocean. Geographic variability was discussed by Tsuda et al. (2000), who reported increases in $E_p$ in regions with strong convection, such as the Maritime Continent in the Western Pacific.

Local enhancements in $E_p$, however, do not necessarily mean correspondingly large momentum fluxes. As discussed by Ern et al. (2004) and Alexander et al.
Gravity Wave Coupling

(2008a), the relation between absolute momentum flux and temperature variance is

\[ |u'w'| = \frac{1}{2} \frac{\lambda_z}{\lambda_h} \frac{g^2 T'^2}{N^2 T_o^2} \]

(3)

where \( \lambda_z \) and \( \lambda_h \) are the GW vertical and horizontal wavelengths, respectively. Comparison of vertical temperature profiles derived from adjacent scans through the atmosphere by the HIRDLS instrument, Alexander et al. (2008a) were able to estimate both \( \lambda_z \) and \( \lambda_h \) and so derive absolute fluxes. This is illustrated in Fig. 5, which shows average values of \( T' \) and \( \rho_o |u'w'| \) for May 2006 (SH early winter). Note not only the relatively high values of \( T' \) over Southeast Asia, but also the correspondingly small flux values. In contrast, there are large temperature perturbations and large fluxes over the southern Andes. These geographic variations in flux can be ascribed to source-related differences in the ratio of \( \lambda_z \) and \( \lambda_h \), which is an important factor in determining GW momentum via Eq. (3). Furthermore, using HIRDLS data Alexander et al. (2008a) were also able to follow waves generated over the southern Andes up to heights in the lower mesosphere.

2.3 In-situ observations

2.3.1 Radiosonde observations

A relatively inexpensive way to derive information on gravity wave activity in the lower stratosphere is to use conventional radiosonde soundings made by national weather agencies (Allen and Vincent, 1995). High vertical resolution (\( \sim 20-50 \) m) soundings are now routinely made and archived by many weather agencies. Analysis of these data enables GW to be resolved, although selection of short vertical wavelength (\( \lambda_z < 7 \) km) waves is favoured (Alexander, 1998).

Wang and Geller (2003) and Wang et al. (2005) exploited routine high-resolution soundings of winds and temperatures made at more than 90 stations to study GW activity in the troposphere and lower stratosphere over the continental USA, Alaska and islands in the Pacific and the Caribbean. Analysis of a 4-year dataset that covers a wide variety of terrain ranging from oceanic to mountainous enabled them to distinguish source from propagation effects. Wang and Geller (2003) reported an overall decrease of stratospheric wave activity with increasing latitude, similar to previous studies (e.g. Allen and Vincent, 1995) in the Australian region. Seasonally, stratospheric wave amplitudes are larger in winter than in summer. Interestingly, tropospheric amplitudes are not always well correlated with stratospheric amplitudes. For example, in the troposphere there are clear GW energy maxima over the Rocky Mountains, whereas in the stratosphere maxima occur over the southeastern United States.

Combining wind and temperature profiles gives propagation directions in both the vertical and horizontal (Vincent et al., 1997). Using the US radiosonde dataset Wang et al. (2005) showed that at least 75\% of wave energy is up going in the lower stratosphere. They also report that the ratio of mean intrinsic frequency to inertial frequency is \( \sim 2.4-3 \), with a weak latitudinal dependence. These results are consistent with previous studies. The dominance of low intrinsic frequency waves is probably
Fig. 6. Average values of momentum fluxes from SPB observations for the Arctic (top) and Antarctic (bottom). Panels a and c refer to zonal fluxes, while panels b and d show meridional values. Note the large fluxes over Greenland and the Antarctic Peninsula.

an observational effect, since the $\sim 5$ m s$^{-1}$ ascent rate of radiosondes is inadequate for sampling long vertical wavelength, high phase speed waves—i.e. this is another manifestation of the “observational selection” issue discussed by Alexander (1998) and Alexander and Barnet (2007).

2.3.2 Superpressure balloon observations

Superpressure balloons (SPB) provide a powerful means for studying GW in the lower stratosphere (Hertzog and Vial, 2001). The envelope of an SPB is inextensible so when it is injected with a fixed amount of gas the pressure inside is greater than ambient pressure at the float level. Hence the balloon floats on a constant density (isopycnic) surface. The advantage of the SPB technique over ground or space-based techniques is that the balloons are advected by the background wind and hence measure the frequency of waves relative to the background flow, the so-called intrinsic frequency, $\hat{\omega}$. It is $\hat{\omega}$ that determines wave properties (e.g. Fritts and Alexander, 2003).

SPB developed by CNES, the French Space Agency, have 8 and 10-m diameters and float at pressure levels between 75 and 55 hPa, corresponding to altitudes of about 17 and 18 km, respectively. The balloons carry a small gondola to measure pressure $p$ and temperature $T$ as well as position using GPS techniques. Early results
were limited by the accuracy of the GPS measurements and by the relatively slow 15-min sampling interval dictated by the rate at which data could be transmitted back to ground stations via the Argos satellite system. More information about the SPB technique and how the data are converted to GW momentum fluxes is given in Vincent et al. (2007), while Boccaletti et al. (2008) provide a detailed error analysis of the retrieved fluxes and phase speeds.

A major SPB campaign was made from McMurdo (78°S) in the SH spring of 2005 during which 27 balloons were launched to study the evolution of the vortex. In preparation for this campaign a number of test flights were made from Kiruna (68°N). Results from Vorcore and the Kiruna test flights enable GW momentum fluxes in the northern and southern polar lower stratospheres to be compared. Figure 6 shows density-weighted ($\rho'_o w'_v$, $\rho'_o v'_w'$) momentum fluxes for flights in the Arctic and Antarctic (Vincent et al., 2007). Note the large flux values in the vicinity of mountain ranges, such as eastern Greenland, the southern Andes and the Antarctic Peninsula. Overall, however, when the fluxes are zonally averaged over regions of high topography and oceanic regions it is found that the fluxes are comparable in magnitude (Hertzog et al., 2008).

Wave amplitudes observed over mountainous regions appear to be significantly more intermittent than values measured over the oceans, with very large fluxes sometimes observed over mountains (Hertzog et al., 2008). Figure 7 illustrates an extreme

![Graph showing vertical displacement and temperature](image)

**Fig. 7.** Vertical displacement (blue, left scale) and temperature (red, right scale) recorded on an 8-m diameter SPB during its passage over the Antarctic Peninsula. The shaded region shows the cross-section through the Peninsula.
case where an 8-m diameter balloon suffered an approximate 1.5 km vertical displacement during a severe mountain wave event over the Antarctic Peninsula. The corresponding peak-to-peak temperature change was about 15 K. From the estimated 1 hr intrinsic period and approximate horizontal and vertical perturbation velocities of 10 m s\(^{-1}\) and 0.7 m s\(^{-1}\), respectively, it is possible to estimate an average momentum flux of 0.7 Pa. This is a factor of about 70 times larger than the average flux in this region. Plougonven \textit{et al.} (2008) provide a more detailed analysis of this event. Such large amplitude waves will break low in the stratosphere, whereas weaker amplitude waves can penetrate well into the middle atmosphere before breaking and depositing their momentum (e.g. Alexander \textit{et al.}, 2008a).

3 Gravity Wave Coupling into the Thermosphere

Large-scale gravity waves associated with magnetic storms are a well-known feature of the ionosphere. After a major storm wave disturbances can produce ionospheric effects on a global scale (e.g. Karpachev \textit{et al.}, 2007). Smaller scale waves propagating from sources such as tropical convection in the lower atmosphere have been suggested as the cause of equatorial spread-\(F\) (e.g. McClure \textit{et al.}, 1998; Prakash, 1999), although Kudeki \textit{et al.} (2007) recently questioned the need for gravity wave seeding of spread-\(F\).

Identifying the sources of waves observed in the thermosphere-ionosphere is a challenging problem. To ray-trace backwards a wave observed in the ionosphere to a lower atmospheric source requires good knowledge of the intermediate wind and temperature fields. These cause the waves to be refracted and advected horizontally. The problem is particularly acute in the region near 100 km where there are often large amplitude tidal wind and temperature fluctuations and very large wind shears (Larsen, 2002). Molecular viscosity and thermal diffusive damping effects also become important in the region above 100 km. This complicates the application of GW ray-tracing techniques.

Recently, Vadas and Fritts (2005) provided an improved form of the GW dispersion relation that incorporates kinematic effects. Applying this new formulation Vadas and Fritts (2004) employed a linear model to study the thermospheric response to a GW spectrum generated by vertical motions induced in a tropospheric mesoscale convective complex. Molecular viscosity and thermal diffusivity act as selective filters on a GW spectrum, allowing only the high-frequency, largest vertical wavelength, waves to propagate to the highest altitudes (Vadas and Fritts, 2005). Strong body forces can be induced by the resulting momentum flux deposition at heights well into the thermosphere. There is also a significant solar cycle effect. GWs penetrate to much higher altitudes during active solar conditions when thermospheric temperatures are high compared with solar minimum conditions (Vadas and Fritts, 2006). This is in part because kinematic viscosity decreases more slowly with altitude when temperature increases, and in part due to the increase in \(\lambda_z\) associated with the higher temperatures.

Vadas (2007) recently explored the properties of GW propagating into the thermosphere-ionosphere in more detail, including the effects of both vertical and horizontal
Fig. 8. Estimated GW spectra as a function of altitude for GWs generated by a deep convective plume and propagating into a thermosphere with an exospheric temperature of 1000 K. A sudden shear of $-100 \, \text{m s}^{-1}$ is assumed at $z = 120$ km. (a) Vertical wavelength spectrum. (b) Horizontal wavelength spectrum. (c) Ground-based wave period spectrum. (d) Ground-based horizontal phase speed spectrum. Shading shows regions of strongest amplitude. (From figure 16, Vadas, Horizontal and vertical propagation and dissipation of gravity waves in the thermosphere from lower atmospheric and thermospheric sources, J. Geophys. Res., 112, A06305, 2007. Copyright 2007 American Geophysical Union. Reproduced by permission of American Geophysical Union.)

dispersion. Figure 8 shows the evolution of GW spectra as a function of altitude for waves generated by a deep convective plume in the lower atmosphere and with a background wind that is westward in the thermosphere. Note how the spectra are dominated by waves with periods $\sim 15$–20 min and that have horizontal scales of about 200–400 km. The largest amplitude waves are propagating eastward, against the mean flow. It should be noted that the effects of ion drag, wave breaking and eddy viscosity are neglected in this calculation (see Vadas (2007) for more details). Despite these limitations the results are in broad agreement with observations.

Sun et al. (2007) also explored the problem of GW propagation into the thermosphere using a three-dimensional transfer function technique. Similar to Vadas (2007) they found that GWs in a 15–30 min period band and horizontal wavelength $\sim 200$–400-km were most likely to propagate to the 300-km level.

While these studies have made significant improvements to our understanding of
GW propagation into the thermosphere-ionosphere system inclusion of the effects of the ionosphere itself are required, since wave motions move charge in the presence of the Earth’s magnetic field. This leads in turn to selective damping, depending on the propagation direction of the waves relative to the magnetic field and the associated ion-drag effects (e.g. Hines, 1968; Tsugawa et al., 2004).

4 Summary and Conclusions

The past decade has seen a great improvement in our understanding of GW coupling and their role in determining the state of the middle and upper atmosphere. Improvements result from the development and application of new observing techniques as well as better appreciation of observational limitations. New satellite and balloon measurements, in particular, provide better understanding of flux variability on both regional and global scales.

In summary:

1. The majority of the waves observed in the lower stratosphere couple energy and momentum upward into the middle and upper atmosphere.

2. Largest momentum fluxes are observed over regions of high topography, but these regions have the greatest wave variability.

3. On a zonally averaged basis, momentum fluxes over mountains and oceans are approximately equal.

4. Thermospheric GWs are selectively filtered by the kinematic dissipation. Only the high frequency, long vertical wavelength components to penetrate to the highest altitudes.

5. There is a strong solar cycle effect in GW propagation into the thermosphere. GWs propagate to higher altitudes during high sunspot conditions than during solar minimum conditions.

Despite the significant progress made in the past decade there are still a number of uncertainties. While it can be relatively straightforward to identify waves associated with topography, it is less easy to identify the GW sources over the ocean. The relative importance of different sources, such as shear and geostrophic adjustment, has still to be established. This is especially true for mid- to high-latitude sources. Convection is the main GW source in the tropics and there are now a number of models that attempt to specify wave fluxes above convection (e.g. Song et al., 2003; Beres et al., 2004; Lane and Moncrieff, 2008), although the model results have yet to be fully tested by observation.

With the deployment of new instruments we can expect good progress in understanding wave coupling into the thermosphere-ionosphere system. New radars, such as the Advanced Modular Incoherent Scatter Radar (AMISR), make three-dimensional simultaneous measurements of ionospheric parameters (Nicolls and Heinselman, 2007). This flexibility provides better insight into the relationship
Gravity Wave Coupling 291

between TIDs and gravity waves and outcomes give hope for better identification of wave sources. In turn, improved appreciation of wave dissipation effects in the thermosphere can now be used to provide better understanding of atmospheric parameters in a region that is otherwise difficult to probe. For example, using GW observations made with AMISR, Vadas and Nicolls (2008) were able to infer the average background horizontal winds in the 160–240 km height range over Poker Flat, Alaska.

The challenge now is to apply these new instrumental developments, as well as advances in theory and modelling, to understand better upward coupling by gravity waves. This will be a major focus of research during CAWSES-II.


References


Hamilton, K., Dynamical coupling of the lower and middle atmosphere: Historical background to current

Hertzog, A. and F. Vial, A study of the dynamics of the equatorial lower atmosphere by use of ultra-long-


