Interpretations of observed climatological patterns in stratospheric gravity wave variance

M. J. Alexander
Department of Atmospheric Sciences, University of Washington, Seattle

Abstract. Observational analyses of gravity waves in the stratosphere have revealed various climatological patterns in gravity wave activity. Seasonal, geographical, and vertical variations have all been observed. In this work, a linear model of gravity wave propagation is applied to investigate the underlying causes of some of the observed patterns. A collection of monochromatic gravity waves that represent a broad spectrum of wavelengths and frequencies is input at 6 km altitude in the model. Propagation of the waves through realistic background atmospheric wind and stability fields is treated with linear ray theory and a simple saturation condition to limit amplitudes to stable values. The wave spectrum at the 6-km source height is specified to be constant at all latitudes, longitudes, and times, so the variability that appears at higher altitudes is due entirely to background atmosphere variations. Before the model results are compared to the observations, the spectrum of waves is filtered in a way that mimics the limitations of each of the observation techniques. The filtering is described in terms of vertical wavelength and is referred to as the "observational filter." In a vertically varying background wind, gravity waves are Doppler-shifted in intrinsic frequency and refracted to different vertical wavelengths as they propagate vertically through the atmosphere. The observational filter and the wave refraction effects can thus couple in interesting ways that have not been explicitly considered in previous work. The model shows that this coupling can give rise to geographical, seasonal, and vertical variations in gravity wave observations without any variations in the spectrum or amplitude of gravity wave sources in the troposphere. Thus careful consideration of both the background wind profile and observational filter can greatly affect the interpretation of the observed climatological patterns in gravity wave activity.

1. Introduction

Gravity waves are mesoscale phenomena with global-scale effects. The momentum they transport vertically to the upper troposphere and middle atmosphere means they can provide an important or even dominant forcing term in the momentum budget at these altitudes. Modelers, however, still lack realistic constraints on gravity wave effects.

Since the import of gravity wave effects has been recognized, many observational studies have been conducted and reported in the literature. A few of these have derived the momentum flux associated with gravity waves which is the important quantity for understanding their effects on the general circulation | Vincent and Reid, 1983, Friis and Vincent, 1987, Friis and Yuan, 1989, Wang and Friis, 1990; Friis et al., 1990; Sato, 1993; Alexander and Pfister, 1995). These studies have, however, been limited to single locations and brief time periods. More often, gravity waves are observed through the perturbations they create in the horizontal wind and temperature fields. From these, estimates of variance are derived and loosely described as measures of gravity wave activity.

There are now emerging some long-term records of gravity wave activity from which seasonal trends in activity can be seen in the stratosphere [Hirota, 1984; Tsuda et al., 1991; Wilson et al., 1991; Allen and Vincent, 1995; Eckermann et al., 1995; Whiteman and Carswell, 1995]. Recent analyses of satellite data have further provided global views of gravity wave activity showing seasonal and vertical variations as well as geographic patterns [Fetterer and Gille, 1994; Wu and Waters, 1996]. A variety of interpretations of the observed variability have emerged. Linear theory predicts exponential growth of wave variance with height. Departures from this are generally interpreted as evidence of dissipation. Seasonal trends have been related to variations in gravity wave sources or source strength, to variable background wind "filtering" (critical level filtering that prevents waves from propagating through the level where their intrinsic frequency goes to zero), and to seasonal variations in density and stability.

Each of the above mechanisms can be important to the interpretation of the observed variability in gravity wave activity. There is, however, another factor that has been largely ignored: Because gravity waves are inherently broadband phenomena, each observation technique imposes a unique observational filter on the total spectrum of waves present, so that only a fraction of the spectrum is visible to the observer. This observational filter selects for waves with certain propagation properties so that when convolved with the effects that "Doppler shifting" or refraction by the background winds can have on the spectrum, it can itself be the cause of the variability in the observations.

In this report, the effects of this observational filtering are quantified and incorporated into a linear model of gravity wave
propagation using a broad spectrum of waves at their altitude of origination. It is assumed that these sources are isotropic in zonal phase speed and perfectly uniform both globally and seasonally. The patterns in gravity wave activity that then emerge from the model are due solely to the effects of background atmosphere variations convolved with the observational filter. Previous authors have taken similar approaches, specifying uniform gravity wave sources in a varying background atmosphere, and have demonstrated some very nonuniform results [Dunkerton and Butchart, 1984; Saravanan, 1990; Eckermann, 1992; Alexander and Rosenlof, 1996, Eckermann and Marks, 1997]. In this paper, model comparisons to wave activity derived from Microwave Limb Sounder (MLS) observations [Wu and Waters, 1996], rocket sounding [Eckermann et al., 1995], and nadir sounder [Allen and Vincent, 1995] analyses will be shown.

The model with uniform gravity waves at their source level will be shown to reproduce a substantial fraction of the observed gravity wave activity variability but not all of this variability. The purpose of exercising this model with uniform gravity wave sources is not to prove that gravity wave sources are uniform but rather to demonstrate background atmosphere effects which must be taken into account before the observations can be interpreted in terms of source variations. The model comparisons to these data further provide some new insights into the physical meaning of the reported averaged variances which have not previously been considered. Quantitative comparisons in these results are still hampered by uncertainties in the shape and strength of the model source spectrum; however, these comparisons are aiding the development of global climatological constraints on gravity wave sources that were discussed by Alexander and Rosenlof [1996].

The results show that the background wind and stability fields have quite important effects on the observations. For this reason, comparisons are here mainly restricted to the stratosphere where we have realistic background atmosphere information. The United Kingdom Meteorological Office (UKMO) assimilated data [Swinbank and O’Neill, 1994] are used. In addition, only zonal winds and zonally propagating gravity waves are considered in the model because the meridional winds are relatively weak in the stratosphere and have only minor effects on the gravity wave spectrum. The linear gravity wave propagation model is first summarized and is then followed by descriptions of gravity wave sources and constraints. In section 3, the observational filters are described, and some key theoretical concepts behind the interpretation of the model results are developed. Patterns in the modeled gravity wave variance are then shown along with descriptions of the observational results. Interpretations of the results, a brief summary, and conclusions follow.

2. Model Description and Constraints

A linear ray model is employed to describe the propagation of a spectrum of waves through realistic background wind and stability fields. Each member of the spectrum is treated as a monochromatic linear wave. The model is very similar to those applied by Eckermann [1992], Marks and Eckermann [1993], Alexander [1996], Alexander and Rosenlof [1996], and Warner and McIntyre [1996] with minor differences in saturation criteria and in the treatment of effects at the highest and lowest frequencies. The largest differences among zonal-mean effects predicted by these models is likely to be in the choice of the gravity wave source spectrum, although no direct comparisons have yet been made. The only change in the present model from Alexander and Rosenlof [1996] is the use of a broader source spectrum, but the resulting zonal-mean effects are not noticeably changed.

The amplitude variation with height is constrained by conservation of wave action flux:

$$F_A = \frac{\overline{\rho E c_{gy}^2}}{\omega} = \text{constant}$$

(1)

where $F_A$ is the vertical component of wave action flux, $E$ is the total wave energy per unit mass, $\overline{\rho}$ is the background density, $c_{gy}$ is the vertical component of wave action flux, and $\omega$ is the intrinsic frequency sometimes called the Doppler-shifted frequency. Considering a zonally propagating wave,

$$\omega = \omega_0 - kU.$$  

(2)

$U$ is the background zonal wind, $k$ is the zonal wavenumber, and $\omega_0$ is the frequency a ground-based observer would measure. Since gravity waves commonly have short horizontal wavelengths (large $k$), and phase speeds within the range of the climatological zonal-mean wind speeds in middle atmosphere, background wind variations $U(z)$ can lead to dramatic variations in $\omega$ and $c_{gy}$ with height.

The applicable dispersion relation includes the effects of rotation at low frequencies and nonhydrostatic effects at higher frequencies (see Appendix A):

$$\omega^2 = \frac{N^2 k^2 + f^2 (m^2 + \alpha^2)}{k^2 + m^2 + \alpha^2}$$

(3)

Here $N$ is the buoyancy frequency, $f$ is the inertial frequency, $m$ is the vertical wavenumber, and $\alpha = 1/(2H)$ where $H$ is the density scale height. This form of the dispersion relation includes the modified high-frequency cutoff discussed by Marks and Eckermann [1995] where total internal reflection of the wave occurs. With this dispersion relation, the vertical group velocity $c_{gv} = d\omega/dm$ is

$$c_{gv} = \frac{-m (\omega^2 - f^2)}{\omega (k^2 + m^2 + \alpha^2)}$$

(4)

If the horizontal displacement of the ray path of the wave is small compared to the scale of the horizontal variations in $U$ then (1) is equivalent to conservation of pseudomomentum flux.

$$F_p = kF_A = \text{constant}$$

(5)

This approximation applies to the model results reported here.

Gravity wave sources are specified in the middle troposphere at 6 km as a two-dimensional spectrum of momentum flux $F_p(k, \omega)$, with $(k, \omega)$ representing a discrete set of waves. The source level background atmosphere is artificially modified to be globally and seasonally uniform in order to ensure truly uniform gravity wave sources but is relaxed back to the observed background state within a few kilometers of the source height, so the stratosphere is unaffected by this lower boundary condition. To model mean flow effects, a convective instability saturation criterion is imposed to limit wave amplitude growth. Wave action flux is conserved along the rays unless the amplitude exceeds the stability criterion. Applying this saturation condition to the wave amplitudes gives
where \( E_w = |w|^2 \) is the square of the vertical velocity amplitude, and the right-hand side is the convective instability limit. \( E_w \) is related to the total energy and wave action flux via the polarization relations [Gossard and Hooke, 1975]. Use of the convective instability criterion may be less appropriate at low frequencies than a dynamical criterion [Lelong and Dunkerton, 1997], but differences in the spectrally integrated solution presented here would likely be small. Multiple breaking levels are accommodated by application of (6) from the lower to the upper boundary. Wherever waves are saturated, they transfer energy and momentum to the mean state. The gravity-wave driven zonal force per unit mass \( F_x \) can be computed as the vertical gradient of the total momentum flux in the spectrum

\[
F_x = \frac{\rho}{\partial z} \sum_{\ell, m} F_{\ell m}(k, \omega, z). \tag{7}
\]

The efficiency factor \( \varepsilon \) is discussed below. Further description and applications of the model can be found in the work of Alexander [1996] and Alexander and Rosenlof [1996].

Gravity wave sources are still quite poorly defined by existing observations. It is very difficult to determine the intrinsic frequency of gravity waves in observations because the ground relative frequency, propagation direction, horizontal wavelength, and horizontal wind must be measured simultaneously, and this has only been accomplished in a few observational studies [Pfister et al., 1993; Swenson et al., 1995; Taylor et al., 1995; Fritts et al., 1997]. To infer properties of gravity wave sources, the waves must also be observed in close proximity to the source because background wind shear between the source and the observation level can substantially modify the observable properties of the waves.

Because of these observational difficulties, the amplitudes and propagation properties of the waves in the source spectrum in the model are very poorly constrained. This gives too wide a range of flexibility in the model results. Constraints on the zonal-mean zonal gravity-wave-driven force at midlatitudes in the mesosphere and stratosphere were developed by Alexander and Rosenlof [1996]. Both theoretical and observational studies [Holton, 1983; Garcia and Solomon, 1985; Fritts and Vincent, 1987; Fritts and Yuan, 1989; Alexander and Rosenlof, 1996] have helped to quantify the role of gravity waves in generating a zonal force in the middle atmosphere that contributes to driving the Lagrangian-mean circulation in the vertical/meridional plane, often called the residual circulation. This circulation with associated tracer transport and adiabatic heating/cooling effects is an important component in determining the zonal-mean structure of the middle atmosphere. The magnitude and latitude/height distribution of the force provide powerful constraints on the source amplitudes and on the efficiency factor [Alexander and Rosenlof, 1996]. The efficiency factor here might best be called an intermittency factor. It is a scaling factor that describes the very real effect of the intermittency of wave sources in space and time.

The shape of the source spectrum \( F_{\ell m}(k, \omega) \) is, however, still poorly constrained, and this uncertainty in shape ultimately results in an order of magnitude uncertainty in the total gravity wave momentum flux at source level [Alexander and Rosenlof, 1996]. Different source spectra also change some of the details of the results at upper levels, but here the focus is on the amplitude and phase of the annual cycle in gravity wave variance, and on trends in latitude. None of these will be greatly affected by the source uncertainties. Only zonally propagating gravity waves are considered here, simplifying the calculation considerably, yet still capturing the dominant effects of the background winds which are primarily zonal in the stratosphere.

The source spectrum employed here includes horizontal wavelengths \( \sim 800, 400, 200, 100, 50, 25, 12.5, 6.25 \) km and periods \( \sim 240, 120, 60, 30, 24, 20, 15 \) min. These give \( 8 \times 14 = 112 \) waves, but the spectrum is further restricted to those with intrinsic phase speeds less than \( 170 \) m s \(^{-1} \), so there are a total of 92 waves used in the calculation. Each wave is assigned the same momentum flux, but with the chosen logarithmic distribution in \( k \) and because the wave energy emphasizes low-frequency waves relative to the momentum flux, the energy spectrum of this source is red in accordance with most observations.

The zonal-mean gravity-wave-driven zonal force \( F_x \) is calculated by zonally integrating the ray paths to produce a single vertical profile of momentum flux and computing the vertical convergence according to (7). In the work of Alexander and Rosenlof [1996], the source spectrum amplitudes and the efficiency factor \( \varepsilon \) were constrained by the following two observed features of the extratropical gravity-wave-driven forcing at the solstices: (1) The sign of the force in the summer stratosphere is negative (westward). The stratospheric force is therefore accelerating, having the same sign as the background winds in the middle and upper stratosphere. (2) The force opposes the wind in the mesosphere (eastward force in summer, westward force in winter), and it peaks with a magnitude of \( \sim 100 \) m s \(^{-1} \) d \(^{-1} \) at mid-latitudes. The expanded spectrum used in this work is constrained in the same manner as in the work of Alexander and Rosenlof [1996], so these results are consistent with this previous work. (See also Appendix B.) In these calculations, \( \varepsilon \) is a simple scaling factor applied as a constant at all latitudes, heights, and times, so it does not affect the seasonal cycle.

3. Model Calculations of Gravity Wave Activity

When the spectrum is calibrated with the above constraints, the total model variance in zonal velocity or temperature can easily be calculated from the same model results. The ray-tracing procedure provides a complete description of the wave propagation properties and amplitudes for each member of the spectrum throughout the model domain.

However, each of the observations to which the model results will be compared measures a uniquely different portion of the spectrum. Thus the total variance in the spectrum must be filtered by an appropriate filter function unique to each observational method. Filter functions appropriate for rocket sounding [Eckermann et al., 1995] and radiosondes [Allen and Vincent, 1995] analyses are shown in Figure 1. Both are band-pass filters. The radiosondes have higher vertical resolution but cover a smaller range of altitudes. The filters are approximate for the long-wavelength cutoff will depend on the details of the shape of the background profiles removed in the original analyses, and these are generally variable. Also plotted in Figure 1 (dashed curve) is the filter function appropriate for the satellite observations by the Microwave Limb Sounder (MLS) [Wu and Waters, 1996] which is essentially a low-pass filter in vertical wavenumber (sensitive only to the very long vertical wavelength waves), so it provides rather different measures of gravity wave activity. The MLS filter is calculated as the normalized power
Figure 1. Observational filter functions showing fractional amplitude observed as a function of vertical wavelength for rocket sounding (solid), radiosonde (dotted), and Microwave Limb Sounder (MLS) (dashed) comparisons. The MLS filter for channels 2/14 is shown, but filters for channels 3-7 and 9-14 are very similar. The MLS mesosphere channel 8 filter has a similar shape but is shifted to the left, toward longer wavelengths. Rocket and radiosonde filters are both modeled as simple band-pass filters with ranges of 2-10 km and 0.125-7 km, respectively.

spectral density function of the vertical weighting function of the MLS measurement. Here the MLS weighting functions are approximated as Gaussian profiles with a full width at half maximum (FWHM) equivalent to the FWHM of the true weighting functions shown by Wu and Waters [1996]. (Note that the MLS limb-scanning observations also filter out horizontal wavelengths greater than ~100 km.) All together, these three observation techniques cover a very broad range of vertical wavelengths.

There is another difficulty in comparing the model and data analyses: that is deciding what the reported observational "average variance" means. The answer depends on the intermittency of the gravity wave events to which each technique is sensitive. Although the model calibration procedure determines a source intermittency \( \varepsilon \), the intermittency of the observed waves will vary as a function of height because the vertical group velocity \( c_g \) will vary substantially via the intrinsic frequency \( \omega \) and vertical wavenumber \( m \) (equations (2)-(4)). High \( \omega \) and low \( m \) go with fast \( c_g \). This is more easily evident in the limit \( N >> \omega >> f \) for which \( c_g \sim \omega m \). One can then think of the residence time \( \tau \) as the time it takes a wave packet to travel a vertical distance \( \Delta z \). Because the group velocity varies with height, \( \tau \) will also be a function of height,

\[
\tau(z) = \frac{\Delta z}{c_g(z)}.
\]

The MLS and sounding observations provide a contrast in this group velocity effect on the intermittency. Figure 2 shows the average \( \tau(z) \), with \( \Delta z = 1 \) km, associated with each of the three observation techniques, which is the average in that portion of the spectrum visible through each of the three observational filters. The MLS, sensitive only to the very long vertical wavelengths, observes only high group velocities. This. Their residence times in the middle atmosphere are very short (\( O \sim 1 \) min/km), and intermittency should be very important. The soundings, in contrast, are only sensitive to short vertical wavelengths which have low group velocities and thus long residence times (\( O \sim 1 \) h/km). It is thus much more likely for the soundings to detect the presence of gravity waves than the MLS at any single observation point in space and time. This conclusion rests on the assumption that there is some intermittency in the gravity wave sources.

These ideas lead to a probabilistic description of the gravity wave variance that can be applied to monthly mean observations. The variance in the observations, \( \nu_0 \), at any given height can be described as the integral over the spectrum \( \nu(k, m) \) of gravity waves multiplied by the observational filter \( \mathcal{F}(m) \) and the probability \( P(k, m) \) that each member of the spectrum is present. For a discrete spectrum the integral is the sum

\[
\nu_0 = \sum_{m, k} P(k, m) \mathcal{F}(m) \nu(k, m).
\]

Here \( \nu(k, m) \) is determined directly from the model. All of the terms in (9) also vary with height. To quantify \( P(k, m) \), information about how the source spectrum is intermittent, both spatially and temporally, is required.

For the purpose of illustration, imagine the components of the gravity wave source spectrum (characterized by \( k, m, \omega \)) turning

Figure 2. Residence times across a 1-km vertical layer as a function of height. (a) Zonal wind for this example. Average residence time for waves observable in the (b) radiosonde analysis of Allen and Vincent [1995], (c) rocket sounding analysis of Eckermann et al. [1995], and (d) MLS analysis of Wu and Waters [1996].
on and off, with the average length of time they are turned on being $t_1(k, m_0)$ and the average length of time between these periods equal to $t_2(k, m_0)$. Then the probability of observing them at some height could be written

$$P(k, m) = r_0(k, m_0) \frac{\tau(k, m)}{t_2(k, m_0)} < 1. \quad (10)$$

Here $\epsilon_0$ is a fractional spatial coverage of waves at the source height. Note that although residence times are quantifiable in the model, they cannot be translated into a probability of observation unless the details of the spatial distributions, time histories of the sources, and their spectral characteristics are all known. The calibration procedure described in section 3 determines only the simple scaling factor $\epsilon$ in equation (1). For the above example, $\epsilon$ would be simply an average value of the spectrum of intermittencies at the source height $\epsilon_0 \cdot t_1(t_1 + t_2)$. In nature, $\epsilon_0$, $t_1$ and $t_2$ could also have seasonal variations. The fractional spatial coverage factor $\epsilon_0$ would be important for satellite observations where the data are averaged over some large area footprint.

This kind of detailed information about sources is not currently known, so we cannot apply equation (9). Instead, consider what patterns in gravity wave variance might emerge in the two extremes of very short and very long residence times. In the case, like the MLS, where the residence times are very short, the patterns are dominated by intermittency because $P(k, m)$ is very small for the observable portion of the spectrum. For the MLS the patterns in an average gravity wave variance show patterns in the probability that any wave will be both present and visible through the MLS filter. The observed variance is thus proportional to the integral over the filtered spectrum

$$\mathcal{V}_{\text{ML}} \propto \epsilon \cdot \sum_{m, k} \mathcal{F}(m) \mathcal{V}(k, m). \quad (11)$$

$\mathcal{F}(m)$ is the filter function in Figure 1, and $\mathcal{V}(k, m) = (\mathcal{T}(k, m)/\mathcal{T})^2$ is the square of the ratio of perturbation to mean temperature. $\mathcal{V}_{\text{ML}}$ is proportional to the likelihood of any wave being present that is visible through the MLS filter.

A second type of "average variance" can be considered when some wave is almost always observed because the observation technique is sensitive to the part of the spectrum with small group velocities and long residence times. In this case, the average variance represents a measure of typical or average wave amplitude in the filtered spectrum

$$\mathcal{V}_R \propto \frac{\sum_{m, k} \mathcal{F}(m) \mathcal{V}(k, m)}{N_s}. \quad (12)$$

This type of average best describes the rocket and radiosonde climatology. For the rocket and radiosonde filters in Figure 1, $N_s = \sum_m \mathcal{F}(m)$ is the total number of observable waves in the spectrum. The two measures of variance in equations (11) and (12) produce fundamentally different geographical and seasonal patterns in "average variance." The proportionality symbols in (11) and (12) are used to emphasize the uncertainty in the absolute value of the variance in these equations caused by this lack of information on gravity wave sources. The right-hand side of equation (12) will give only an upper limit for the observed variance because the intermittency is assumed to be unity here when integrated over the spectrum.

### 4. Results

#### 4.1. MLS Comparison

Wu and Waters [1996] (hereinafter referred to as WW96) analyzed radiances from saturated channels during the MLS routine limb-scanning observations. They report monthly mean fractional temperature variances representing horizontal perturbations with scales of $\sim 10-100$ km for January and July 1993. Global maps of these variances at seven different log-pressure levels (Plate 1) show that peak variance occurs in winter at mid-latitudes within the stratospheric westward jet. Weaker summer hemisphere maxima occur in the 5°-25° latitude band with centers of enhanced variance over Madagascar and Brazil in January and over southeastern Asia and the Gulf of Mexico in July. The magnitudes of these summer and winter maxima become more similar with increasing altitude in these observations. The tropical region between 5° in the summer to 30° in the winter hemisphere contains relatively very low variance.

WW96 propose that the winter midlatitude variance maximum is coincident with the strongest winds because upstream propagating waves "grow more efficiently with height" in those regions. They also suggest that variations in gravity wave sources play an important role. Specifically, convection over tropical and subtropical landmasses and flow over topography are discussed as important causes of the geographic variations in the MLS variance.

Results of the ray model and equation (11) are shown in Plate 2 for comparison to the MLS analysis of WW96. To create the maps in Plate 2, the isotropic, constant gravity wave source is applied at each point $(x, y, z_0, t)$ with $z_0 = 6$ km and $t$ being January and July 1993. Monthly mean UKMO-assimilated winds and temperatures [Swinbank and O'Neill, 1994] are used to describe the background state. Only waves with $k \geq 2\pi/(100)$ km are included, approximating the horizontal wavelength filtering of the WW96 observations, but this filtering in $k$ tends only to reduce the magnitude of the variance and has negligible effect on the geographical, seasonal, and height variations. The horizontal displacement of the ray paths has a negligible effect on these solutions because the observable waves have such high intrinsic frequencies that they propagate very nearly straight up. Equation (11) is applied at each height as the spectrum evolves in response to the zonal wind and stability variations. A weighted vertical average is then computed by using Gaussian approximations to the weighting functions of the five MLS 63-MHz channels in the stratosphere (WW96). (Note that no comparison in the mesosphere is possible with the UKMO background state because these data only extend to the 0.3-hPa level.) The value of $\epsilon$ applied in equation (11) is the one constrained by the gravity-wave-driven force in equation (7), although it should be remembered that this scaling factor is still somewhat arbitrary for these variance calculations as discussed above.

The global patterns and seasonal and vertical variations in MLS variance are reproduced quite well with this model without any global or seasonal variations in sources. In particular, the summer/winter contrasts and latitudinal structure are very similar, with maxima occurring in the winter jet stream and a secondary maximum in the summer subtropics in a patchy band between $\sim 5^\circ$ and $25^\circ$ latitude. Centers of activity consistently appear in the subtropical band over Madagascar and Brazil in January, although this longitudinal variability does not appear to be so pronounced as it is in the WW96 observations.

The geographic patterns in the modeled variance maps in Plate 2 result from specifying uniform gravity wave sources, so
source variations play no role. The strength of the zonal winds is an important factor in determining the patterns in Plate 2 but not because wave amplitude growth is more efficient in some regions than in others. Instead, it is only in the regions of strong background winds that much wave activity occurs with a sufficiently long vertical wavelength to survive the severe filtering associated with the MLS observation technique. (See Figure 1.) Only upstream propagating waves with large intrinsic frequencies and phase speeds have these long vertical wavelengths. This model and the interpretation also explain why the variance at the equator is so low despite plenty of convection for forcing waves at these latitudes. It should be noted that the MLS observations are more sensitive to meridionally propagating gravity waves than those propagating zonally at low latitudes (WW96). Since meridionally propagating waves were not considered in the model, this omission might account for some of the more pronounced longitudinal variations seen in the data in the $5^\circ$-$25^\circ$ bands.

WW96 also plot vertical profiles of zonally averaged variance for winter latitudes $50^\circ$-$70^\circ$ and summer latitudes $5^\circ$-$25^\circ$. Their results (reproduced in Figure 3) show exponential growth in variance, proportional to $e^{-\sqrt{\gamma}H}$ in the stratosphere, and a transition to nearly zero growth in the mesosphere, beginning at about 50-km altitude. WW96 suggest the transition is caused by wave breaking and that the altitude of the transition shows where gravity wave drag occurs. They further suggest that this low 50-km altitude of transition suggests "the stratosphere jets would be closed at somewhat lower altitudes" at these latitudes.

Figure 4 shows the model results for comparison. The model (Figure 4a) gives the same growth rate in the stratosphere, and it also captures the transition to slower growth beginning at about 50 km in the upper stratosphere. To examine the behavior of these profiles in the mesosphere, a calculation using COSPAR International Reference Atmosphere (CIRA) [Fleming et al., 1990] to specify the background atmosphere is shown in Figure 4b. CIRA winds are meant only to represent some mean climatology and are quite unrealistic at low latitudes, so the comparison to WW96 is not so good. However, the profiles do show the transition to zero growth in the mesosphere. In the model the transition from exponential growth to zero growth in variance does not result from wave breaking in the upper stratosphere. Wave breaking is instead occurring at all levels, and the peak gravity wave drag in the mesosphere occurs about 20 km above the altitude where the curves in Figure 4 become constant. The transition to zero growth is directly related to the observational filter imposed by the MLS observation technique rather than to wave dissipation. Without filtering, the total variance in the model grows steadily with height throughout the range 30-80 km as $-e^{-\sqrt{\gamma}H}$. The total variance is much larger than the filtered variance at all levels and grows more slowly because critical-level filtering and dissipation occur at all levels. With the observational filter, only the variance associated with the very longest vertical wavelength waves contributes. Figure 4c shows the zonal mean winds in the $50^\circ$-$70^\circ$ latitude bands for the two winter seasons. In the stratosphere the winds increase steadily with height, so gravity waves propagating upstream, against the wind, are "Doppler shifted" or refracted to long vertical wavelengths that can be seen through the MLS filter. (See also the schematic in Figure 5.) At $-50$-$60$ km the background wind shear changes sign. Above this level, those upstream propagating waves now begin to be refracted back toward smaller vertical wavelengths and out of the region of the spectrum visible to the MLS. Also above this level, the other half of the spectrum that is propagating eastward begins now to be refracted toward longer vertical wavelengths. However, these waves have been severely attenuated by critical-level filtering and dissipation in the stratosphere below, so they do not carry so much energy. Thus Doppler shifting by the background winds, the MLS vertical wavelength filtering, as well as critical-level filtering all conspire to produce the shape of the profiles in Figure 4. Since the MLS observation technique is sensitive only to a portion of the total spectrum, it would be a mistake to extrapolate the "saturation" of the observed variance profiles to the energy growth of the full gravity wave spectrum likely to be present. Fritts and Lu [1993], for example, summarize previous observations which place the altitude of the transition to zero growth of the gravity wave energy profile much higher, near 85-100 km.

The main conclusion of this model comparison is that the patterns in gravity wave variance that the MLS observes result primarily from a convolution of the spectral Doppler shifting (or refraction) caused by the background vertical shear with the vertical wavelength filtering effects of the observation technique itself. Geographic distributions of sources (if they play some role in shaping the patterns in the observations) cannot be discerned from the MLS observations until these effects of the background winds are carefully considered.

4.2 Rocket Sounding Comparison

Eckermann et al. [1995] analyzed 11 years (1977-1987) of rocket soundings from 14 launch sites in the northern hemisphere at latitudes ranging from $9^\circ$ to $77^\circ$N and one site at $8^\circ$S. There were no observations between $38^\circ$ and $55^\circ$N. The data were analyzed in two height ranges, 20 40 km and 40 60 km and sorted by month of the year to infer the seasonal cycle of gravity wave variance at each launch site. Their results (reproduced in Figure 6) show a large annual cycle at high latitudes ($55^\circ$-$77^\circ$N) with maximum in winter and minimum in summer. At latitudes lower than $-35^\circ$ a peak in August appears. At still lower latitudes, $-8^\circ$-$20^\circ$, this peak broadens to include the summer months June to August.

The model comparison to the rocket sounding analysis of Eckermann et al. [1995] is shown in Figures 7 and 8. Figure 4 shows latitude/time maps of fractional temperature variance.
Plate 1. Maps of observed gravity wave radiance variances from the analysis reported by Wu and Waters [1996] of the MLS observations. The variances are normalized by the squared mean radiance brightness temperature and plotted on a logarithmic scale. Wind vectors are also overlapped. Taken from Wu and Waters [1996].

Figure 4. Modeled vertical profiles of zonal mean variance using the MLS filter in four latitude bands using (a) United Kingdom Meteorological Office (UKMO) data and (b) COSPAR International Reference Atmosphere (CIRA) to define the background atmosphere. Comparing these to the observations (Figure 3), the same exponential growth with height is seen in the stratosphere as well as the transition to constant variance in the mesosphere. Latitudinal and seasonal variations in (Figure 4a) are also very similar to the observations. CIRA represents a climatology and is unrealistic at the low latitudes but provides a reasonable comparison at the high latitudes in the mesosphere. (c) Zonal mean zonal wind for 50°-70° N January (solid) and 50°-70° S July (dotted) from CIRA. See the discussion in the text.
Plate 2. Modeled maps of gravity wave temperature variance calculated with the linear ray model, constant, isotropic and uniform gravity wave sources, and equation (11) with the MLS filter. Compare these to the MLS gravity wave observations in Plate 1. Maps corresponding to the five MLS 63-MHz channels in the stratosphere are shown for both (a) January and (b) July 1993 monthly means.

($T'/T)^2$ in units of $%^2$ averaged over two height regions, 20-39 km and 40-56 km, similar to Eckermann et al. [1995]. To create these maps, the same isotropic and constant source spectrum at $z_0 = 6$ km is applied at each latitude and month. The rocket filter in Figure 1 is then applied in equation (12) to compute the average variance at each height as the spectrum evolves, propagating through monthly and zonal mean UKMO data. Altitude in this case is the geometric height above the surface, so the UKMO data have been redefined in these coordinates. This distinction between log-pressure altitude and geometric altitude turns out to be important to the interpretation of the results and will be discussed below. Results from 4 years, 1992-1995, are averaged together to show the seasonal cycle. Note that equation (12) contains no efficiency factor but rather describes only average variance. Despite the long residence times (Figure 2), some residual intermittency is likely to be present. So the variance in Figures 7 and 8 has been scaled by a factor 0.1 for easier comparison to the observations and to represent this residual intermittency.

The model results in Figures 7 show the same strong annual cycle at high latitudes with peak variance in the winter months. At low latitudes in the model, a July to August peak dominates, just as was observed in the data. The strength of the modeled annual cycle is similar to the observations but is weaker at the high latitudes. These same features all appeared as the dominant patterns in the observational results. The model's seasonal patterns in zonal wind variance (not shown) are very similar to those in Figure 7. Hirota [1984] also compiled a map of gravity wave activity from a smaller set of rocket soundings (Hirota's Figure 7) which looks very similar to Figure 7.

The model produces considerable interannual variability shown in Figure 8. Figure 8a shows the temperature variance at 21°N in both altitude regions for 1992 through 1995. The August peak at the upper level stands out clearly in this time series but is superimposed on considerable interannual variability resulting entirely from the background atmosphere variations. Note the large February peak in 1992 which does not appear in other years. This sensitivity of the model to the background state can help explain some of the details of the differences between the model and the observations. Similar interannual variations were also seen in the data [see Eckermann et al., 1995, Figure 12].

The August peak in the model (Figure 8a) results primarily from seasonal wind variations. Figure 8b shows the zonal mean winds as a function of time and height at 21°N. In August there is weak monotonic shear throughout the upper troposphere and stratosphere up to about 40 km. This promotes weak but steady refraction of waves toward the smaller vertical wavelengths. In September these positive effects of refraction only extend to about
25-30 km. Because the winds never reach really large magnitudes in August, critical level filtering is also not too severe. In June the shear extends to higher altitudes, leading to more severe critical level filtering of the observable portion of the spectrum. So conditions of weak monotonous shear and moderate maximum wind speeds are favorable for observing waves in the range of vertical wavelengths within the rocket sounding band-pass filter (Figure 1).

Figure 8c shows the time variation of the model variance at 79°N. Here the strong winter peaks stand out clearly. These peaks result from the strong annual cycle in density on a constant geometric height surface. Eckermann [1995] described a simple model of this effect in which the variance is proportional to \((N^2/N_0)(p_0/p)\), where the zero subscript refers to the source level. This factor is plotted for 79°N in Figure 8d. The density factor in this expression dominates the seasonal cycle when this factor is computed at a constant geometric altitude. Stability and wind variations play a secondary role but contribute significantly to the interannual variability. At the low latitudes, conversely, seasonal variations in winds play the dominant role. Eckermann's [1995] model did not include effects of background winds so did not reproduce the observed July to August peak at latitudes \(10°-35°\), although background wind effects were discussed as likely to be important in explaining the data.

Vertical profiles of both wind and temperature variance (not shown) have both similarities to and differences from the observations. Both data and model show similar profile shapes. The model uses zonal mean winds 1992-1995, while the rockets were launched at specific geographic sites between 1977 and 1987. It is therefore not surprising that comparing data and model at specific latitudes and months reveal different profile shapes. The UKMO winds are also quite uncertain in the upper stratosphere. Consistent with the observations (and with Eckermann's [1995] conceptual model), this model tends to show weaker growth in temperature variance than in horizontal wind variance in the upper stratosphere above \(\sim 35\) km.

4.3 Radiosonde Comparison

The model can also be compared to radiosonde analyses. Allen and Vincent [1995] show seasonal gravity wave activity patterns in the Australian sector of the southern hemisphere between June 1991 and May 1992. Their results (reproduced in Figure 9) show an annual cycle in gravity wave energy density at midlatitudes, \(35°-42.5°\)S, with maximum in winter and minimum in summer. At low latitudes, \(12°-22°\)S, a strong maximum is observed in summer during the wet season, November to March. The highest energy densities were observed in the tropics. These radiosonde results represent averages over the log-pressure-altitude range \(\sim 17-24\) km.

The sensitivity of the model results to the background wind means that a good comparison cannot be derived for 1991 since the UKMO winds are not available for most of this period. Interannual variability is especially important at tropical latitudes where the quasi-biennial oscillation dominates the lower-stratosphere wind variations. The seasonal cycle of radiosonde-filtered wave activity is shown for the 4-year period 1992-1995 in Figure 10. Temperature variance computed with equation (12) is converted to energy density with the formula given by Allen and Vincent. The model has been scaled by the factor 0.1 as for the rocket sounding results for easier comparison to the observations and to account for some intermittency. In Figure 10, the 44°S model shows a similar annual cycle as in the radiosonde results, with peak activity in winter (July). The pattern at lower latitudes, however, does not much resemble the observations. In particular, there is no peak in activity during the tropical wet season (December to May) evident in the model results shown for 11°S in Figure 7. The assimilated UKMO winds are less accurate at low latitudes which might account for some of the discrepancy, but a seasonal cycle in the gravity wave sources may be the more likely explanation at the low latitudes.

The midlatitude winter peak in variance is not associated with the seasonal cycle in density as it was in the case of the rocket soundings. Instead, these peaks coincide with latitudes and times where the wind profiles show weak, nearly monotonous shear, a mechanism similar to that creating the low-latitude August peak in the rocket sounding data.

The close proximity of the radiosonde observations to the likely wave sources in the troposphere may make the radiosonde observations more sensitive to the properties of the gravity wave sources [Hamilton and Vincent, 1995; Allen and Vincent, 1995; Nasstrom et al., 1997; Vincent et al., 1997] than the MLS or rocket-sounding results in the upper stratosphere. Detailed comparisons of these data and model results will be the subject of future work.

5. Summary

A linear ray model with saturation was used to describe climatological patterns in observed gravity wave activity. The model was constrained to start with a globally uniform, zonally isotropic, and constant gravity wave source spectrum specified in the middle troposphere. This spectrum evolves with height in response to variations in the background wind and stability. This same model was also applied by Alexander and Rosenlof [1996] and Ray et al. [1997] to examine seasonal patterns in the gravity-wave-driven zonal forcing in the stratosphere.
To compare the modeled wave activity to the data, observational filters were constructed (Figure 1). These describe how the vertical resolution and analysis methods limit which portions of the total gravity wave spectrum are visible in the observational results. Model comparisons to MLS (WW96), rocket sounding [Eckermann et al., 1995], and radiosonde [Allen and Vincent, 1995] analyses are shown. The model describes the dominant patterns seen in both the MLS and the rocket data very well. High sensitivity of the model to the background atmospheric state can produce considerable interannual variability. The comparison to the radiosonde observations suggests that these lower-altitude observations may retain more memory of their sources.

The physical explanations behind the patterns in variance seen in the MLS and sounding observations are quite different. Equation (11) describes the MLS variance as proportional to the unlikely probability of observing the long vertical wavelength waves in which the MLS is sensitive. The long vertical wavelength waves have high group velocities, so they do not spend much time in the middle atmosphere, passing through rather quickly (Figure 2d). The sounding observations, conversely, are best described by equation (12) which says the variance is proportional to the most typical amplitude of the waves observable in the soundings. These waves have shorter vertical wavelengths, much slower vertical group velocities, and reside in the middle atmosphere much longer (Figures 2b, 2c). Thus when the observation is made, it is probable some waves will very often be present, and
The variance patterns match patterns in average amplitude. These differences in vertical wavelength and group velocity are also connected to other properties of the waves such as the dominant propagation direction of the observable waves. In the case of MLS the waves propagating upstream are mainly observed [Wu and Waters, 1997], and this result also appears in the model, following directly from application of the MLS observational filter (Figure 1).

6. Concluding Remarks

The three sets of observational results discussed here show similar annual cycles at middle to high latitudes with enhanced wave activity in the wintertime and relatively low activity during summer, and this is a common feature of gravity wave observations. It is tempting to suggest that the enhancement is somehow related to enhanced forcing in the strong winter winds, or due to reduced filtering of stationary waves during winter. Allen and Vincent [1995] proposed winter storm fronts as a seasonally varying source responsible for the seasonal cycle in wave activity in the midlatitude radiosonde observations. Although many of all these mechanisms may be operating, none of these effects need be considered to explain the observed seasonal cycles in the three data sets discussed in this paper. Although wind filtering sometimes causes the seasonal patterns in the model, the source spectrum is a broad one with no momentum flux at zero ground-relative phase speed, the traditionally assumed phase speed for orographic waves. The assumed broadband source spectrum, when atmospherically and observationally filtered, reproduces many aspects of patterns of observed gravity wave climatologies. Different source spectra (when appropriately filtered) may, in fact, give similar levels of agreement. However, these different source spectra can predict substantially different vertical and horizontal distributions of gravity wave mean flow forcing in the middle atmosphere.

Despite the fact that all three observations show similar annual cycles at middle to high latitudes, the model results suggest that the mechanism responsible is unique to each. In the case of the rocket-sounding results the seasonal variation in atmospheric density on a constant geometric height surface is the controlling mechanism. Density variations, however, have little influence on the winter maximum observed by the MLS and radiosonde techniques which correspond to log-pressure-altitude coordinates. For MLS it is the zonal wind variations that dominate the patterns. Only where the wind speeds are large does the long-wavelength portion of the spectrum contain much energy. The differences in the underlying mechanism causing the seasonal cycle depend directly on the different observational filters associated with each technique. As can be seen in Figure 1, there is little overlap in the vertical wavelength domain between the MLS and the sounding filters. These two types of observations then are looking at two unique portions of the spectrum, so the waves that each technique observes have unique propagation properties and fundamentally different interactions with the atmosphere. This underscores what is perhaps the obvious fact that no single observation technique can see the whole spectrum of waves likely to be present but also suggests something less obvious: that extrapolating the results of a single set of observations to the effects of the full spectrum of waves can lead to erroneous conclusions.

An analogy best illustrates this last point. Consider the vertical wavelength filter as the view of an observer seen from the outside. A packet of wave activity can pass in and out of this narrow field of view by Doppler shifting/refraction effects. The change in the wave packet’s position relative to the window frame can be thought of as the change in vertical wavelength of the wave packet as a function of altitude. In interpreting observations, it must be remembered that a packet of wave activity can be refracted to a region of the spectrum outside of the observational filter without any dissipation occurring.

The same argument applies to the momentum flux carried by the waves. If the appearance or disappearance of momentum flux within the observer’s window is taken literally without considering this larger context, simple refraction effects could be interpreted as convergence of momentum flux and could lead to erroneous conclusions about the wave-driven mean flow forcing.

The influences of Doppler-shifting on the frequency spectrum of ground-based observations were explored in some detail by Friis and VanZandt [1987]. The differences between observed and intrinsic frequencies are now routinely considered in the observational literature. The influence of the background winds in shaping results observed at different vertical wavelengths seems, however, to be unrecognized or largely ignored. For both these reasons, observations of gravity wave activity, to be interpreted meaningfully, must consider the local profiles of back-
ground atmospheric stability and winds. The results shown here demonstrate that the observed variability in the background state can produce large variations in observable gravity wave activity without any variations in their sources. Because planetary-scale waves can act as a variable background state to the smaller-scale gravity waves, they too will need to be quantified if gravity wave variability is to be studied on shorter timescales than the monthly means discussed here. Other climatologies of gravity wave activity exist for higher altitudes [Vincent and Fritts, 1987; Senft and Gardner, 1991; Nakamura et al., 1996; Thomas et al., 1996]. The same model applied here can be applied at these higher levels as realistic background atmospheric fields become available.

It should also be emphasized that quantitative comparisons between model and data are hindered by a lack of knowledge of the characteristics of wave sources and their intermittency. Fourier spectral analysis gives average wave characteristics over data analyzed in this way. This and other techniques that average many observations eliminate information on intermittency origi-

Figure 9. Time-latitude contours of total gravity wave energy density derived from radiosondes over Australia. The energy densities are averaged over the altitude range 17-24 km and shown for the period June 1991 to May 1992. Taken from Allen and Vincent [1995].
density varying only in \(z\), \(f\) is the inertial frequency, \(g\) the gravitational acceleration, and \(c_s\) is the sound speed. \(N\) is the buoyancy frequency

\[
N^2 = -g \left( \frac{1}{\rho} \frac{\partial \bar{p}}{\partial z} + \frac{g}{c_s^2} \right)
\]

and the total derivative is

\[
\frac{D}{Dt} = \frac{\partial}{\partial t} + U \frac{\partial}{\partial x} + V \frac{\partial}{\partial y},
\]

where \((U, V)\) is the background wind. For a zonally propagating wave with wavenumber \((k, m)\) propagating in the \((x, z)\) plane, the solutions are

\[
(u', v', w') = \tilde{p}_i^{-1/2} \exp \left[ \frac{1}{2} \int (kx + mz - \omega_n^2) \right],
\]

\[
(p', \rho') = \tilde{p}_i^{+1/2} \exp \left[ \frac{1}{2} \int (kx + mz - \omega_n^2) \right].
\]

Substitution of these into (A1)-(A5) with the definition of density scale height \(H = -\tilde{p}_i (\partial \tilde{p} / \partial z)^{-1}\) gives the dispersion relation, equation (3).

Appendix B: Correction to Alexander and Rosenlof [1996]

The units given for integrated momentum fluxes in the gravity wave source spectra were incorrectly given as \(\text{kg m}^{-1} \cdot \text{s}^{-2}\) in the text on pp. 23,468-23,469. The correct units are \(\text{m}^2 \cdot \text{s}^{-2}\). To convert to the former, these must be multiplied by the atmospheric density at 6 km altitude which is approximately 0.6 kg m\(^{-3}\).

Acknowledgments. This research was supported by the Physical Meteorology Program of the National Science Foundation, ATM-952546, and the National Aeronautics and Space Administration, NASA grants NAG-11803 and NAGS 3188. Steve Eckerman, James Holton, Charles McLandress, Leonard Pfister, Kaoru Sato, and the anonymous reviewers provided valuable comments on the manuscript. Dong Wu provided preprint versions of his color figures for reproduction.

References


M. J. Alexander, Department of Atmospheric Sciences, University of Washington, Box 351640, Seattle, WA 98195-1640. (e-mail: alexand@atmos.washington.edu)

(Received April 16, 1997, revised September 23, 1997, accepted November 11, 1997.)