QUARTERLY JOURNAL OF THE ROYAL METEOROLOGICAL SOCIETY

Q. J. R. Meteorol. Soc. **00**: 1–35 (2009)

Published online in Wiley InterScience

(www.interscience.wiley.com) DOI: 10.1002/qj.000

Recent Developments in Gravity Wave Effects in Climate Models, and the Global Distribution of Gravity Wave Momentum Flux from Observations and Models

M. J. Alexander[1], M. Geller[2], C. McLandress[3], S. Polavarapu[4], P. Preusse[5], F. Sassi[6], K. Sato[7], S. Eckermann[6], M. Ern[5], A. Hertzog[8], Y. Kawatani[9], M. Pulido[3], T. Shaw[10], M. Sigmond[3], R. Vincent[11], S. Watanabe[9]

[1]*NWRA/Colorado Research Assoc., Boulder, CO, USA,* [2]*SUNY-Stonybrook, New York, USA,* [3]*Dept. of Physics, Univ. of Toronto, Toronto, Ontario, Canada* [4]*Environment Canada,* [5]*Forschungzentrum Jülich, Germany,* [6]*Naval Research Laboratory, Washington, DC, USA,* [7]*Univ. of Tokyo,* [8]*UPMC Univ Paris 06, Laboratoire de Météorologie Dynamique, France,* [9]*Japan Agency for Marine-Earth Science and Technology, Yokohama, Japan,* [10]*NYU,* [11]*Physics, Univ. of Adelaide, Australia*

Abstract: Recent observational and theoretical studies of the global properties of small-scale atmospheric gravity waves have highlighted the global effects of these waves on the circulation from the surface to the middle atmosphere. The effects of gravity waves on the large-scale circulation have long been treated via parametrizations in both climate and weather forecasting applications. In these parametrizations, key parameters describe the global distributions of gravity wave momentum flux, wavelengths, and frequencies of the waves. Until recently, global observations could not define the needed parameters because the waves are small in scale and intermittent in occurrence. Recent satellite and other global data sets with improved resolution along with innovative analysis methods are now providing constraints for the parametrizations that can improve the treatment of these waves in climate prediction models. Research using very high resolution global models has also recently demonstrated the capability of resolving gravity waves and their circulation effects, and when tested against observations, these models are showing some very realistic properties. Here we review recent studies on gravity wave effects in stratosphere-resolving climate models, recent observations and analysis methods that reveal global patterns in gravity wave momentum fluxes, and results of the very high resolution model studies, and we outline some future research needs to improve the treatment of these waves in climate simulations.

KEY WORDS class file; LaTeX 2$_\varepsilon$; *Q. J. R. Meteorol. Soc.*

Received July 2009; Revised ; Accepted

1 Introduction

Small-scale atmospheric waves, called gravity waves (or internal waves), have sources in the troposphere such

*Correspondence to: M. J. Alexander, NWRA/Colorado Research Associates Division, 3380 Mitchell Lane, Boulder, CO, 80301, USA

as flow over topography, convection, and jet imbalance. As these waves propagate upward and dissipate, they force the atmospheric circulation at altitudes near the tropopause and well above, in the stratosphere and mesosphere. Global circulation models used for weather and climate forecasting may resolve some gravity waves but include the effects of unresolved gravity wave forcing via parametrizations. The parametrizations compute a momentum forcing term by making assumptions about the unresolved wave properties that have not been well-constrained by observations. The assumptions are formulated as a set of tuning parameters that are used to adjust the circulation and temperature structure in the upper troposphere and middle atmosphere.

Different models have different goals and apply different types of parametrizations with unique tunings. For example, chemistry-climate models use gravity wave parametrizations to adjust their stratospheric circulation and polar temperatures to realistic values, important for accurate simulation of ozone chemistry. Climate and weather forecasting models have traditionally used mountain wave parametrizations to improve the structure of the winter jets and horizontal temperature gradients near the tropopause, and improve surface wind distributions. Recent studies have highlighted additional important effects of gravity waves in climate models that underscore the need to reduce uncertainties in gravity wave parametrizations. The requirements needed from observations to constrain the tuning parameters in gravity wave parametrization schemes have been impossible to achieve on the global scale until fairly recently.

What is needed is characterization of the vertical flux of horizontal pseudomomentum, and three-dimensional (3D) wave propagation properties. We will use the common approximation for pseudomomentum flux as momentum flux, which for gravity waves can be written as the product of density and the covariance of horizontal u' and vertical w' wind perturbations. It is often written as $\bar{\rho}(\overline{u'w'}, \overline{v'w'})$, the overbar represents a spatial or temporal average over a wavelength or period. The 3D wave properties include horizontal and vertical wavelength, as well as propagation direction. Only in the last decade have high-resolution satellite observations allowed the estimation of these wave properties and momentum fluxes. However, each individual satellite measurement technique can only observe a portion of the full 3D spectrum of gravity waves. Errors in the momentum fluxes derived from global observations remain large, although local case studies can now be quite accurate. Some very high resolution global modelling studies that include a middle atmosphere are now resolving much more of the gravity wave spectrum and are being used to study the wave sources, propagation and dissipation, and the momentum forcing of the circulation without the use of any gravity wave parametrizations. In recent years, new methods for inferring the missing momentum forcing due to unresolved waves in lower resolution models have also been developed using advanced data assimilation methods.

With these recent developments, we can collectively better constrain and quantify the role of gravity waves in the momentum budget of the global circulation. The purpose of this review is to provide an overview of recent developments in this rapidly developing field.

2 Gravity wave issues in stratosphere-resolving global climate models

General circulation models used for climate studies of the troposphere-stratosphere system have generally used relatively coarse resolution ($\sim$ 2–5° in the horizontal and variable in the vertical but usually $\sim$ 3km in the stratosphere) and prescribed sea surface temperatures (SSTs). Depending upon the problem at hand, they may be either dynamics-only models (e.g., those that use prescribed ozone fields to compute solar heating) or models coupled to a fully interactive chemistry scheme (e.g., those that use predicted ozone to compute solar heating). Given that multi-year simulations are necessary, their coarse resolution is dictated by current computional resources, a situation that is not likely to change much in the foreseeable future.

While their coarse resolution is generally fine enough to capture most of the Rossby wave activity in the atmosphere, it is not fine enough to resolve the bulk of the small-scale gravity waves that play a large role in the momentum budget of the middle atmosphere. To account for the effects of these unresolved disturbances, gravity wave drag (GWD) parameterzations must be used. Following the pioneering work of Lindzen (1981) and subsequent refinements by Holton (1982), the first implementation of GWD parametrizations in global climate models was in the context of orographically-forced gravity waves (Palmer et al., 1986; McFarlane, 1987). However, as model lid heights increased into the mesosphere, parametrizations for non-orographic gravity waves (e.g., gravity waves forced by convection and other sources) became necessary. As a result, GWD parametrizations are generally grouped into two categories: 1) orographic GWD (OGWD) schemes, which represent zero phase speed waves forced by subgridscale topography, and 2) non-orographic GWD (NGWD) schemes, which represent non-zero phase speed waves forced by mechanisms other than flow over topography. See McLandress (1998) for a review of a number of those parametrizations and their impact in climate model simulations.

In order to make the GWD parametrization problem tractable and computationally efficient, a number of important simplifying assumptions must be made about the waves and their sources, including the spectrum of momentum flux. Assumptions common to current parametrizations are that the gravity waves propagate only vertically and instantly through the column, and that the waves propagate in a conservative fashion up to a height at which an amplitude threshold is exceeded, the wave breaks, deposits momentum flux, and generates GWD (see references in McLandress, 1998). All GWD parametrizations employ tunable parameters which act to scale the wave drag and/or change the breaking heights. In principle, these parameters should be constrained by observations. In practice, however, they are determined by tuning the models to observed climatologies.

In the following three subsections, we discuss a number of recent issues concerning the impact of parameterized GWD on global simulations of climate and climate change. We focus on the region below the stratopause since that is where these recent advances have occurred.

2.1 Issues regarding angular momentum conservation

The transfer of momentum by gravity waves is an important process in the global angular momentum budget of the atmosphere. Parametrizations of this transfer must respect the principle of conservation of angular momentum, which represents an important constraint on the dynamics. Since current GWD parametrizations assume that momentum transfers are strictly vertical, angular

momentum conservation is equivalent to zonal momentum conservation. Shepherd and Shaw (2004) formulated a momentum conservation constraint for GWD parameterization in a model vertical column, which requires that any gravity wave momentum flux across a vertical model level must be entirely absorbed in the atmosphere above. The constraint expresses the fact that the momentum budget is closed with regard to the upper boundary (i.e., there can be no radiation of gravity wave momentum flux to space).

The implications of the gravity wave momentum conservation constraint for modelled climate have been explored by Shaw and Shepherd (2007) and Shaw et al. (2009). Using a zonally-symmetric mechanistic model, Shaw and Shepherd (2007) showed that violating momentum conservation leads to a non-robustness in the response of parametrized nonorographic gravity waves to a radiative perturbation in the middle atmosphere. In particular, when momentum was not conserved, the response became sensitive to changes in the gravity wave source spectrum, parametrization scheme, background flow and model lid height. Momentum conservation was violated by either letting gravity wave momentum flux escape to space or through the use of a Rayleigh friction layer. The sensitivity to model lid height resulting from nonconservation explained the sensitivity to lid height found by Lawrence (1997) and is unacceptable for model intercomparison purposes. The spurious response resulting from nonconservation was also non-negligible in terms of its magnitude and downward influence into the troposphere. The downward influence is described by the "downward control" mechanism (Haynes et al., 1991), whereby wave drag in the stratosphere induces meridional circulation cells that extend into the troposphere and modify temperature gradients, zonal wind patterns, and sea-level pressure. When momentum was conserved the response became robust to changes in the above mentioned factors.

Using a comprehensive global climate model, Shaw et al. (2009) quantified the impact of conserving both orographic and nonorographic gravity wave momentum flux on 3D modelled climate and on the climate response to idealized ozone depletion. When the model lid was placed at 10 hPa, nonconservation led to large biases in the mean climate and its variability, which extended to the surface. The errors due to nonconservation resulted not only from the direct effect associated with the missing force attributed to the missing gravity wave momentum flux but also from errors in the resolved planetary waves in the model. The effects on the planetary waves resulted from a feedback through the modelled winds, which were always closer to reality when momentum was conserved. In practice the errors which arise from nonconservation could be corrected by re-tuning the gravity wave parametrization parameters, however this would be tuning against an error. When the model was perturbed by idealized ozone depletion there were large biases, both in the magnitude and in the sign of the planetary wave and parametrized gravity wave responses relative to previous modelling studies (Manzini et al., 2003). When the model lid was placed at 0.001 hPa, there was little impact of momentum conservation on both the mean climate and its response to the idealized ozone depletion, which is expected due to the small amount of gravity wave momentum flux reaching 0.001 hPa.

Momentum conservation in GWD parametrization is an important theoretical constraint. It is most useful when the total amount of gravity wave momentum flux input into the parametrization is accurate. Observations of gravity wave momentum flux are key to constraining GWD

parametrizations and in conjunction with momentum conservation ensure the physical realism of the large-scale response to the GWD.

2.2 Issues regarding orographic GWD

The need to parametrize the effects of small-scale gravity waves generated by unresolved topography in global models has been acknowledged for over two decades (e.g., Palmer et al., 1986; McFarlane, 1987). Inclusion of their effects leads to substantial improvements in simulations of the mid-latitude lower stratosphere in northern winter, and, by remote influence, the troposphere. This occurs as follows: The bulk of the OGWD occurs in the mid-latitude lower stratosphere, where it leads to a deceleration of the zonal wind in that region. These weakened winds cause a separation of the subtropical and polar night jets, which in turn alters the refractive properties for planetary waves, leading to changes in the patterns of resolved wave drag (Sigmond and Scinocca, 2010). The impact of both the parametrized and resolved wave drag on the troposphere then follows from "downward control" (Haynes et al., 1991). A number of recent studies indicate that the "downward control" signal in the troposphere can be amplified by feedback mechanisms involving tropospheric eddies (e.g. Song and Robinson, 2004; Chen and Zorita-Gotor, 2008).

Thus, in order to get the stratospheric and tropospheric responses to OGWD right for the right reasons, it is important that the tunable parameters be constrained by observations. A common tunable parameter in OGWD schemes is a so-called efficiency factor (ϵ), which acts to scale the drag without affecting breaking heights. The Whole Atmosphere Community Climate Model (WACCM; Garcia et al. 2007) uses a variant of the McFarlane (1987) scheme, and calls the efficiency factor e. In the Scinocca and McFarlane (2000) scheme, the efficiency factor is included in the parameter G. The values of these parameters are typically chosen to be those which produce the smallest biases in quantities such as zonal mean zonal winds and mean sea-level pressure (e.g., Palmer et al., 1986, McFarlane, 1987, Scinocca et al., 2008). Separate parameters can change the breaking heights (e.g., "$\mathrm{Fr}_{\mathrm{crit}}$ in Scinocca and McFarlane, 2000). Observations are not likely to be able to directly constrain such parameters, but can constrain physical quantities such as momentum flux and intermittency. (See Section 3.) From parametrizations in models, it is possible to reconstruct and output these physical quantities so that in future studies, the tuning parameters can be constrained by observations through this indirect route.

Figure 1 shows climatological zonal mean zonal winds for December–February from the Canadian Centre for Climate Modelling and Analysis (CCCma) model (which uses the Scinocca and McFarlane (2000) scheme), computed using two values of G (Sigmond and Scinocca, 2010). The "weak" setting (left) produces more realistic zonal mean winds in the lower stratosphere, but at the expense of less realistic mean sea-level pressures (not shown). The "strong" setting (middle) produces overly weak winds in the lower stratosphere, but more realistic mean sea-level pressures. The wind differences (right) are, in fact, accentuated when the model's upper boundary is raised to the mesosphere (results not shown).

In addition to having a strong impact on climatological mean quantities, parametrized OGWD also has a strong impact on variability, particularly in the stratosphere (Siskind et al., 2007; Richter et al., 2010). Figure 2 illustrates the annual cycle of interannual standard deviation of temperature at 80N and 30 hPa calculated from

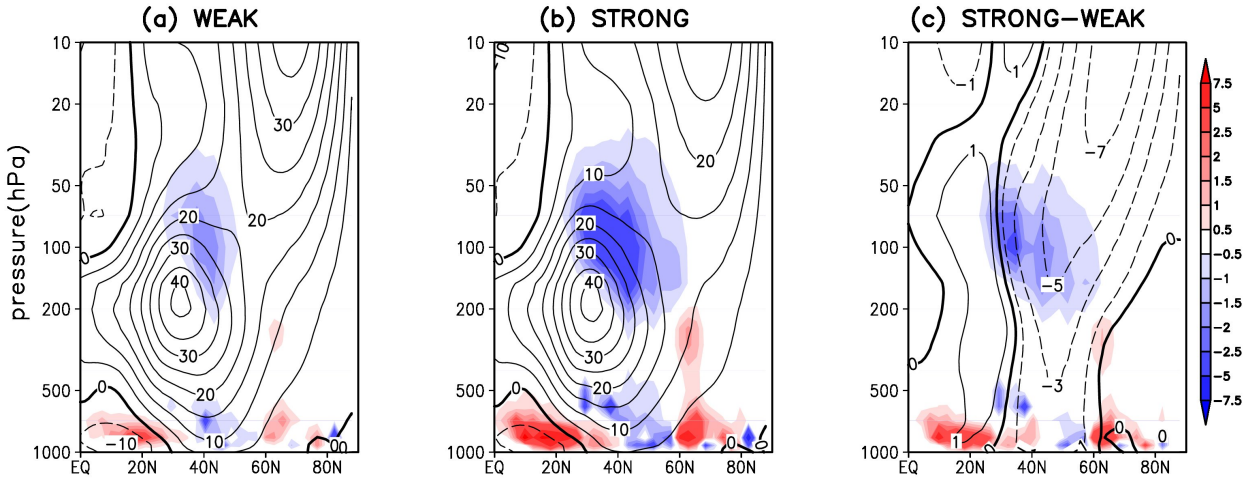

Figure 1. Zonal-mean zonal wind (contours) and OGWD (shading, in 10^{-6} Pa m^{-1}) in NH winter (DJF) for control simulations with a) weak drag (G=0.25) and b) strong drag (G=1.0), and c) its difference. (After Sigmond and Scinocca (2010)).

daily data. The ERA40 reanalysis (red) shows large variability during northern winter peaking around the middle of February, comparatively smaller variability during summer and a progressive increase during fall into early winter. Results from two simulations of WACCM forced with observed SSTs are also shown: In the first simulation (dash-dot), a value is used for the efficiency factor ϵ that produces the best tropospheric simulation (Garcia et al., 2007); in the second simulation (solid black), ϵ is reduced by a factor of 1.6^{-1}. Overall, both simulations are deficient compared to the reanalysis when contrasting the standard deviations during winter. This deficiency leads to a reduced number of stratospheric warmings (Richter et al., 2010). However, the simulation with the smaller value of ϵ (solid) is much closer to the reanalysis and produces more realistic statistics of stratospheric warmings (not shown). The improved variability in this second model simulation is due to small but important changes of the zonal mean zonal wind in the lower stratosphere, which favor a more poleward concentration of planetary wave activity in the winter-time polar vortex (not shown).

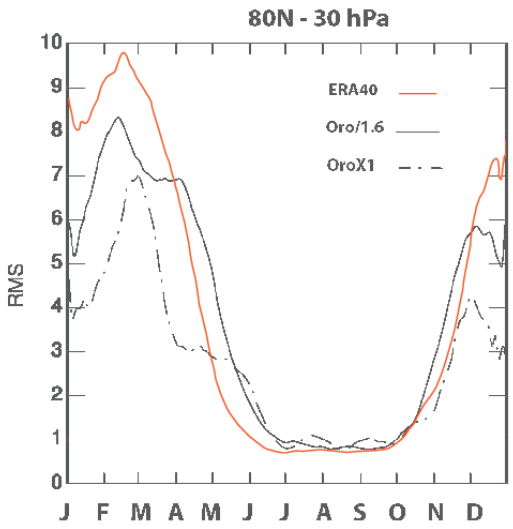

Figure 2. Annual cycle of northern hemisphere interannual temperature variability determined from daily data from ERA40 (red) and two WACCM simulations with differing values of the OGWD efficiency factor ϵ.

Given the above sensitivity of the current climate to these parameters, it is important to quantify their impact on climate change simulations. Sigmond et al. (2007) showed that in the GFDL AM2 model the 500 hPa circulation response at high latitudes depends strongly on the tunable parameter G. In another study, Sigmond et al. (2008) showed that the Northern Annular Mode (NAM)

response to a CO_2 doubling perturbation is more dependent on G than on the model lid height in the CCCma model. Their result left open the question of whether the NAM response to CO_2 doubling (and associated regional climate responses) could depend directly on the poorly constrained parametrized OGWD response. Alternatively, the NAM response may instead be mainly sensitive to the model basic state winds, which the OGWD parametrization settings help to shape. Sigmond and Scinocca (2010) addressed this directly with controlled model experiments that separately doubled CO_2 and varied the OGWD. They found that the basic state winds were the controlling factor shaping the climate response to CO_2 doubling, whereas the parametrized OGWD response had very little direct effect. The important effect of the OGWD is to minimize the biases in the lower stratospheric basic state and to set the refractive properties for resolved waves.

OGWD has a strong and robust impact on the climate-change induced increase in the equator-to-pole Brewer-Dobson circulation in the stratosphere in models (Brewer, 1949) as inferred from Butchart et al. (2006) and confirmed subsequently by Li et al. (2008), McLandress and Shepherd (2009), and Butchart et al. (2010). The mechanism, which is quite straightforward, is as follows: Increased greenhouse gases warm the troposphere and cool the stratosphere. This increases the meridional temperature gradient in the subtropical middle and upper troposphere, which, through thermal wind balance, increases the strength of westerlies on the upper flank of the subtropical jet. Since a large fraction of the parametrized orographic waves occur over the Himalayas, the drag that these waves produce occurs as a result of wave breaking above the subtropical jet maximum. As climate change increases the mean wind speeds in this region, the parametrized waves break higher, causing increased wave drag in the lower stratosphere. (The above argument implicitly assumes that OGWD reacts passively to changes in the zonal wind brought about by climate change: Sigmond and Scinocca (2010) demonstrated that this is, in fact, the case.) Through "downward control" this causes increased upwelling on the equatorward side of the OGWD maximum.

2.3 Issues regarding the ozone hole

Because parametrized OGWD depends on geographic variations in resolved mesoscale topography, the role of OGWD is considerably weaker in the Southern Hemisphere than in the Northern Hemisphere. Consequently, NGWD has played a larger relative role in controlling modelled winds and temperatures in the winter and spring in the Antarctic. Garcia and Boville (1994) showed that mesospheric NGWD can significantly impact the circulation and the thermal structure in the stratosphere through downward control. They showed that the effect on the temperature and vertical residual velocity is largest in the upper stratosphere and decreases at lower levels. Although the magnitudes of the anomalies associated with NGWD are small when compared to the mean climatology, chemical processes are particularly sensitive to slight changes of the ambient temperature and even differences of a few degrees can have a large impact on the activation of the heterogeneous reactions. In this respect, the timing of those changes is even more important. In fact, Garcia and Boville (1994) show that near 30 km the temperature difference in July at 80S can be as large as 10 K, when comparing simulations with and without NGWD; the same temperature difference in September is smaller but is still several degrees (see their Fig. 5). Heterogeneous processes critical for ozone depletion are sensitive to temperatures in the lower stratosphere throughout winter and early spring (Solomon, 1988).

Tuning of NGWD parametrizations with interactive chemistry presents challenges that were unforeseen several years ago. Without interactive chemistry, tuning was only required for the climatological mean temperatures and zonal mean zonal winds, while now with interactive chemisty, tuning for the seasonal cycle is also required. Figure 3 illustrates the difficulties that can be encountered with interactive chemistry. Two simulations using WACCM (Garcia et al., 2007) with 1995 boundary conditions for chemical composition are shown for the southern hemisphere. The two simulations differ only in the amount of non-orographic gravity wave momentum flux at the source level. The simulation with the lower value (0.7 mPa; left panels) shows a more realistic southern hemisphere ozone minimum. By contrast, the simulation with a larger source level momentum flux (1.0 mPa; right panels) shows a much weaker, shorter duration ozone hole.

The remarkable difference of column ozone between the two simulations of Figure 3 is related to seasonal temperature variations in the lower stratosphere which directly impact ozone depletion chemistry. The lower panels of Figure 3 illustrate the annual march of lower stratospheric temperature at 61 hPa over the south polar cap in these two simulations. With larger source level momentum flux (lower right) a more rapid increase of temperature occurs during southern spring compared to consistently lower temperatures between September and November in the simulation with smaller source level momentum flux (lower left). Temperatures in the left panel remain too cold in December, while on the right, the warming in the lower stratosphere occurs too soon. The October monthly mean temperature is already above 200 K in the simulation with larger source momentum flux, sufficient to inhibit the formation of polar stratospheric clouds and the chemical conversion of chlorine into active forms that deplete ozone in early spring. The two settings of source level momentum flux differ by only a small amount relative to the uncertainties in these fluxes.

3 Global observations of gravity wave momentum fluxes

The above model issues call for improved constraints on momentum fluxes in the lower stratosphere. The first direct observation of momentum fluxes (in the mesosphere), $\mathbf{M} = (M_x, M_y) = \bar{\rho}(\overline{u'w'}, \overline{v'w'})$, by gravity waves was made by Vincent and Reid (1983) using a split beam technique with an HF radar. This technique has since been used in the stratosphere at VHF radar sites in Japan, Peru, the UK, and India beginning in the early 1990's (e. g. Sato, 1993; Riggin et al., 1997). In recent years though, observational studies using several other techniques are now providing global-scale information on gravity wave momentum fluxes in the stratosphere. A common thread in these observations is that they analyze fluctuations about some mean as an indicator of gravity waves and then estimate the associated gravity wave amplitudes, and use gravity wave polarization relations to derive the momentum fluxes either from temperature measurements only or from wind and temperature measurements. Other than radar and superpressure balloon studies, all the other techniques discussed in this paper use spatial fluctuations to derive gravity wave momentum fluxes, and all techniques have inherent horizontal and vertical resolutions. As we will see, these spatial resolution limits also imply measurements of certain gravity wave frequency ranges.

3.1 The gravity wave dispersion relation

Gravity waves are treated here and in the analyses cited in the paper as small-amplitude perturbations to some larger

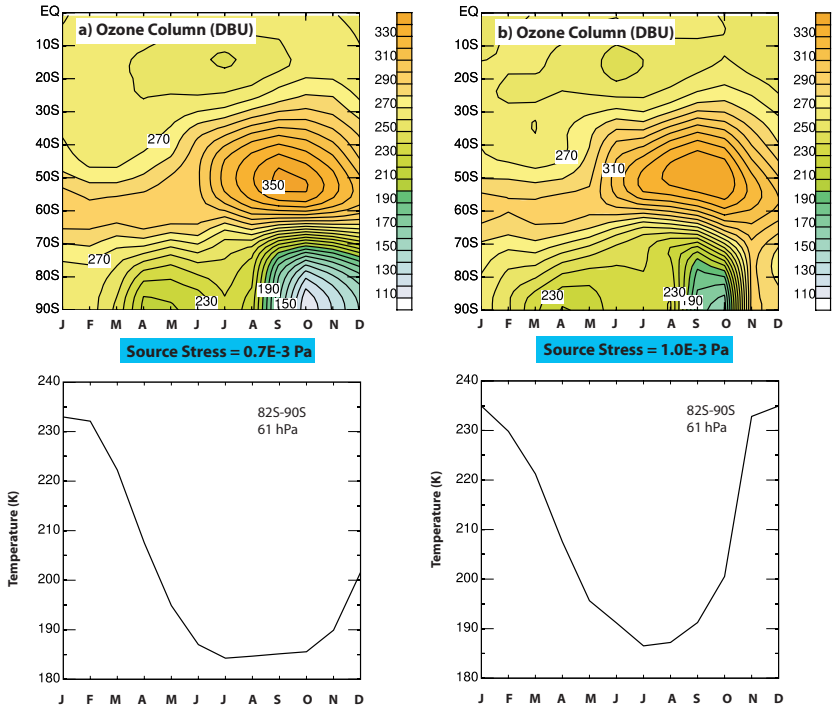

Figure 3. Top panels: Time-latitude changes in southern hemisphere column ozone in WACCM simulations with slightly lower (left, 0.7 mPa) and higher (right, 1.0 mPa) values of NGWD source-level momentum flux. Bottom panels: Time series of polar temperatures at the 61 hPa pressure level in the simulations.

scale horizontally uniform and steady background state with horizontal wind $\bar{u}, \bar{v}$, temperature $\bar{T}$ and density $\bar{\rho}$. We refer the reader to Fritts and Alexander (2003; section 2) for a derivation of the wave equations starting from the fundamental fluid equations, which can be used to derive the dispersion relation.

The dispersion relation relates the wave frequency ω to other wave propagation properties such as the horizontal (k, l) and vertical (m) wavenumbers. An important wave frequency that emerges is the intrinsic frequency $\hat{\omega} = \omega - k\bar{u} - l\bar{v}$, that which would be measured in the frame of reference moving with the background wind. The dispersion relation can be generally written (e.g. Fritts and Alexander (2003); their eqn. (23)) as

$$\hat{\omega}^2 = \frac{N^2(k^2 + l^2) + f^2[m^2 + (2H)^{-2}]}{k^2 + l^2 + m^2 + (2H)^{-2}} \qquad (1)$$

where N is the Brunt-Vaisälä frequency; f is the Coriolis parameter; and H is the density scale height. (We are further neglecting compressibility effects that lead to acoustic-gravity waves.) With slow variations in height $N(z)$, $H(z)$, $\bar{u}(z), \bar{v}(z)$, (1) allows variations $m(z)$ and $\hat{\omega}(z)$. Altitudes where $\hat{\omega} \to 0$ (or $c = \omega/k = U$, where U is the wind component in the direction of wave propagation) are called critical levels. Near or below this level the wave generally becomes unstable and will dissipate. These processes underlie the concept of "critical level filtering".

Various observation techniques have different inherent resolutions, and the method of separating waves from the larger-scale background can further restrict the range of wavelengths that are resolved in an analysis. A restricted range of wavelengths also implies a restricted range of frequencies through (1), a point made earlier by Alexander (1998). We will describe the limitations of

individual techniques in sections 3.3-3.5 and compare the various observable ranges of wavelength and frequency in section 3.7.

3.2 Calculation of gravity wave momentum fluxes

The method of calculating gravity wave momentum fluxes depends on which variables are being directly observed. For instance, the primary satellite-observed quantity is temperature derived from radiances. In this case, the gravity wave polarization relations can be used to derive the implied horizontal and vertical wind fluctuations, and these, in turn, can be used to derive gravity wave momentum flux using the following formula from Ern et al. (2004):

$$(M_x, M_y) = \frac{1}{2}\bar{\rho}\left(\frac{g}{N}\right)^2\left(\frac{\hat{T}}{T}\right)^2\left(\frac{k}{m}, \frac{l}{m}\right) \qquad (2)$$

where $(k,l) = (2\pi/\lambda_x, 2\pi/\lambda_y)$ is the horizontal wavenumber vector, $m = 2\pi/\lambda_z$ is the vertical wavenumber, and $\hat{T}$ is the temperature amplitude of the wave, and T a background (or larger-scale mean) temperature. Equation (2) uses a simplified form of the dispersion relation appropriate for medium frequency waves, $\hat{\omega} = N|(k,l)/m|$. Radiosonde techniques, instead keep terms that are important at low frequencies. Appropriate assumptions are given in the sections that follow.

3.3 Radiosondes

Standard radiosonde data are meant to supply temperature, wind, and humidity data for weather prediction models, so information on only a small number of mandatory and significant levels (*e.g.* Eskridge et al., 1995) are saved and sent to weather prediction centers. These data cannot easily be used to obtain information on gravity waves. Allen and Vincent (1995) illustrated how gravity wave information could be obtained using all measurements in the original radiosonde profiles. This motivated an effort by the World Climate Research Program's (WCRPs) Stratospheric Processes and their Role in Climate (SPARC) project to save these high vertical-resolution radiosonde data (Hamilton and Vincent, 1995). For instance, there are presently nine years of US high vertical-resolution (nominally 30 m vertical resolution) archived on the SPARC Data Center (see http://www.sparc.sunysb.edu/html/hres.html).

Radiosonde data contain information on the horizontal and vertical wind components as well as temperature, but Geller and Gong (2009) illustrated how different gravity wave frequencies are preferentially observed in these different variables. For this reason, gravity wave momentum fluxes are derived from the high vertical-resolution radiosonde data using the wind and temperature measurements together with the gravity wave polarization relations and the assumption that wave energy flux is upward by the following formulae,

$$(M_x, M_y) = -\bar{\rho}\frac{\hat{\omega}g}{N^2}(\overline{u'\hat{T}_{+90}}, \overline{v'\hat{T}_{+90}})\delta_-(\hat{\omega}) \qquad (3)$$

where $\delta_-(\hat{\omega}) = (1 - f^2/\hat{\omega}^2)$ and $\hat{T}_{+90}$ is determined by Hilbert transforming $\hat{T}$ by 90° and then taking the reverse transform (see Vincent et al., 1997). Using the US high vertical-resolution radiosonde data, Gong et al. (2008) show time series of the vertical fluxes of zonal and meridional momentum by gravity waves at individual radiosonde launch sites. These results averaged in longitude over all sites are shown in Figure 4.

The zonal momentum fluxes tend to be positive (eastward) at low latitudes, negative (westward) at middle latitudes, and positive during the summer months at high latitudes. This pattern is broadly consistent with what would be expected from critical level filtering by the background winds: There are primarily westward winds

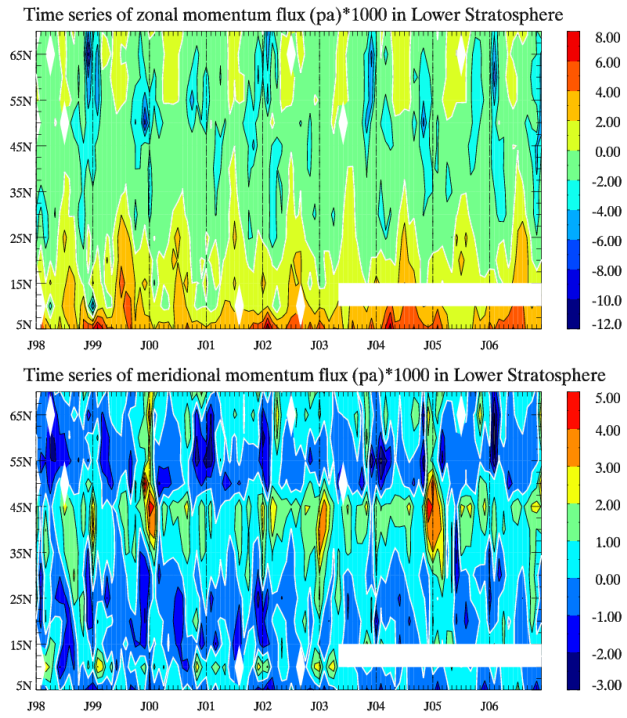

Figure 4. Time-latitude plots of monthly mean eastward (top) and northward (bottom) momentum fluxes derived from US high vertical-resolution radiosondes in units of 10^{-3} Pa for the lower stratosphere (18-25 km). The 9-yr time series at each latitude is an average of sites in that latitude bin. Locations of the sondes are shown in Gong et al. (2008). These are extensions of earlier results by Wang (2003).

at the low latitudes in the upper troposphere/lowermost stratosphere, and the westward winds extend furthest poleward in mid-summer. At middle latitudes the eastward jet stream winds dominate the filtering, giving rise to the negative zonal momentum fluxes there. The picture for the meridional momentum fluxes is less coherent, but generally, the momentum fluxes are equatorward at low latitudes, poleward in middle latitudes, and equatorward again at high latitudes, possibly a signature of the three-celled meridional circulation of the troposphere. As indicated earlier, these radiosonde-derived momentum fluxes are for gravity waves with low intrinsic frequencies, commonly $\hat{\omega} \sim 2f - 4f$, and horizontal wavelengths ∼500-2000 km.

3.3.1 *Long-duration balloons*

Superpressure balloons are closed and made of an almost inextensible material, so that they maintain a fixed constant volume during their flights provided the internal gas pressure exceeds that of the atmosphere. Hence these balloons drift on constant-density (isopycnic) surfaces in the atmosphere, and are capable of long-duration flights (up to several months) that cover wide geographical areas (e.g. The TWERLE Team, 1977; Hertzog et al., 2007). Observations of meteorological variables made at regular time intervals along the flight permit detection of disturbances associated with gravity waves. A unique feature of long-duration balloon observations is that they are collected in the intrinsic reference frame: Balloons are advected by the horizontal wind, and therefore behave as quasi-Lagrangian tracers of horizontal motion so that wave intrinsic frequencies are directly measured.

Gravity-wave momentum flux calculations based on long-duration balloon observations were pioneered by Massman (1981) using results from a few balloons launched during the Tropical Wind, Energy Conversion, and Reference Level Experiment (TWERLE) in 1975-76. Massman's results were, however, confined to short periods of time because of the limited data-transmission capabilities available at that time. These limitations are no longer an issue because of advances in space-borne communications and positioning systems. Estimations of gravity-wave momentum fluxes representative of equatorial latitudes were derived with 3 long-duration balloons flying in the lower stratosphere (around 60 hPa) in 1998 (Hertzog and Vial, 2001). Recently, global momentum-flux maps were obtained at the same level over the poles of both hemispheres with the dataset collected in the framework of the Stratéole/Vorcore project (Vincent et al., 2007;

Hertzog et al., 2008). In these latter studies, GW momentum fluxes are computed from the correlation between the horizontal velocity and atmospheric pressure fluctuations. The pressure disturbance measured by the balloon is linked to the wave-induced vertical velocity disturbance (Boccara et al., 2008). It was also assumed that the balloons behave as perfect isopycnic tracers, which is a realistic approximation for the waves examined in these studies but may be not applicable for shorter period waves (Nastrom, 1980) that will be resolved in future campaigns. For Stratéole/Vorcore, the data rate was limited to (15 min)$^{-1}$, so waves with intrinsic periods $\geq$1 hr were resolved. The range of wavelengths that can be analyzed with the existing data will be further described in section 3.7 and illustrated graphically in Figure 8 in context with other measurement techniques. The hydrostatic approximation to the observational limit of these data gives $k \leq m/12$.

Results of the Vorcore campaign show zonal-mean high-latitude momentum fluxes $\sim 2-3$ mPa, but larger average values near 30 mPa in a $10^\circ \times 5^\circ$ region in the vicinity of the Antarctic Peninsula. The fluxes are largely westward, but significant localized meridional fluxes also occur over topography. The intermittency in these measured fluxes will be described in section 3.6.

3.4 Nadir scanning satellites

For the past 30 years, nadir-viewing passive microwave and infrared remote sensors have been a cornerstone of operational numerical weather prediction, particularly thermal O_2 and CO_2 channels from which temperatures can be derived. However, until recently, two properties of these measurements made them incapable of resolving gravity waves. First, the nadir view yields broad weighting functions in the vertical, such that vertically short gravity-wave oscillations are averaged out in these nadir radiances. Second, early instruments had horizontal resolutions of several hundred kilometers, making them unable to resolve typical gravity wave horizontal wavelength scales of ~5-500 km. For example, the cross-track scanning Microwave Sounding Unit (MSU), which flew on the NOAA-6 through NOAA-14 satellites, had horizontal measurement footprints in excess of 100 km.

Starting with the NOAA-15 satellite, launched in 1998, MSU was superseded by the Advanced MSU Unit-A (AMSU-A), whose footprint diameters were reduced to ~40 km at nadir and whose radiometric precision was improved to ~0.2-0.8 K in the middle atmosphere. Wu (2004) isolated fluctuations in along-track AMSU-A radiances that revealed enhancements well above nominal noise floors near the vortex edge and over orography that correlated with similar gravity-wave-induced radiance enhancements observed by the Upper Atmosphere Research Satellite–Microwave Limb Sounder (UARS-MLS). The launch of NASA's Aqua satellite in 2002 saw AMSU-A integrated with the Atmospheric Infrared Sounder (AIRS) to form a high-resolution coupled temperature sounding system (Parkinson, 2003). In addition to many more thermal channels for analysis, the horizontal footprint for the AIRS measurements are three times smaller than the corresponding AMSU-A measurements, while channel noise levels are comparable. This improved horizontal resolution makes AIRS sensitive to gravity waves of shorter horizontal wavelength than AMSU-A.

AMSU-A and AIRS are not purely nadir viewers: The instruments scan cross-track symmetrically about nadir. When coupled with the satellite motion, a two-dimensional radiance image with a cross-track swath width of ~2000 km is assembled. Wu and Zhang (2004) presented initial evidence that horizontal wave structure

was captured in AMSU imagery. To investigate this, Eckermann and Wu (2006), using a simple in-orbit forward model of lower stratospheric (channel 9) radiance acquisition, showed that two-dimensional horizontal phase structure of large-amplitude long-wavelength gravity waves could be resolved (e.g., horizontal wavelength and propagation direction). Eckermann et al. (2006) validated the model predictions for a large-amplitude mountain wave over Scandinavia, reproducing the observed AMSU-A radiance perturbations via forward modelling of high-resolution numerical weather prediction model simulations of the event. Eckermann et al. (2007) extended the analysis of that wave to all AMSU-A stratospheric channels 9–14, which revealed the wave propagating right through the stratosphere and evolving in phase structure with height due to directional wind shear.

Alexander and Barnet (2007) showed large-amplitude mountain wave oscillations imaged in AIRS swath radiances from an upper-stratospheric 15 μm CO_2 channel. Subsequent studies of radiances from various stratospheric AIRS channels have revealed large-amplitude stratospheric mountain waves over various mountainous regions (e.g., Alexander and Teitelbaum, 2007; Eckermann et al., 2007; Limpasuvan et al., 2007; Eckermann et al., 2009; Alexander et al., 2009) as well as stratospheric gravity waves generated by convection (Grimsdell et al., 2009). Gravity waves imaged in AIRS 4.3μm thermal channels validate earlier imaging of apparent upper-stratospheric gravity wave structures in these band radiances with the SPIRIT-3 radiometer (e.g., Picard et al., 1998; Dewan et al., 1998).

To date, studies have focused primarily on radiances, rather than temperature retrievals, due to the closer connection between the measured perturbation and the properties of the gravity wave that produced it (e.g., Eckermann and Wu, 2006). Nonetheless, an accurate wave-induced temperature amplitude is required to deduce more derived wave properties, such as momentum flux. One approach is to derive first the wave's vertical and horizontal wavelengths from the radiance data, then use those wavelengths to quantify the vertical and horizontal smearing of the wave due to the vertical weighting functions and horizontal footprints, respectively, and then use that information to correct the radiance perturbation amplitude for this smearing to derive the wave's true temperature amplitude. This process works well in case studies (e.g., Eckermann et al., 2006; Alexander and Teitelbaum, 2007; Eckermann et al., 2009; Alexander et al., 2009), but must be done individually and carefully for each event in question. At lower altitudes where AIRS radiances become affected by clouds, this fails. Here, a detailed AIRS/AMSU-A temperature retrieval algorithm is needed in which AMSU-A radiances are used to "cloud-clear" AIRS radiances prior to retrieving temperatures (Susskind et al., 2006). These AIRS/AMSU-A temperature retrievals sometimes retain gravity wave structures seen in the original radiances (e.g., Alexander and Teitelbaum, 2007; Eckermann et al., 2007) but at the expense of desampling to the coarser AMSU-A footprint resolution. Whether gravity wave vertical structure is retrieved optimally by these algorithms is not currently clear. Recently, Hoffmann and Alexander (2009) developed an AIRS-only temperature retrieval. While this retrieval preserves the fine horizontal resolution of the original AIRS data, it has higher noise and does not extend to lower altitudes like the AIRS/AMSU-A temperature retrievals. Nonetheless, for short deep gravity waves that are believed to carry much of the gravity-wave

momentum flux, the Hoffmann-Alexander AIRS temperature retrievals look to be an exciting new product to better quantify the gravity wave momentum budget globally in the middle atmosphere.

Momentum fluxes have been derived in the winter stratosphere for mountain wave events observed over Scandinavia, the Antarctic Peninsula, and South Georgia Island with AMSU and AIRS using equation (2). Peak values in the Scandinavian event were 300 mPa (Eckermann et al., 2007) and for the Antarctic event 140 mPa (Alexander and Teitelbaum, 2007). These were wave events with horizontal wavelengths of 400 and 300 km respectively. Events over S. Georgia described in Alexander et al. (2009) included a range of horizontal wavelengths from ∼50-400 km. Averaged over a $2.5^\circ \times 2.5^\circ$ area, the S. Georgia momentum fluxes ranged 60-200 mPa, while local values were >1000 mPa. Despite these large fluxes, the island is too small to be identified as a land point in global climate models, and the parameterized OGWD is zero. Fluxes from islands in the southern oceans may be an important missing source of wave drag for the southern hemisphere circulation.

3.5 Limb scanning satellites

Infrared limb sounding was the first technique used to detect GWs from space (Fetzer and Gille, 1994). In this geometry the instrument views slightly downward towards the Earth's horizon, through the atmosphere and into cold space (Preusse et al., 2002). The instrument measures the infrared radiation from thermal emissions along the line of sight (LOS). Most of the radiance stems from the lowest part of the LOS close to the tangent height, where the LOS is almost parallel to the Earth's surface. Therefore infrared limb sounding is sensitive to GWs with vertical wavelengths longer than 2-5 km (depending on instrument), and primarily sensitive to GWs of horizontal wavelengths longer than 100-200 km (Preusse et al., 2002). Altitude profiles are taken by varying the viewing angle and thus the tangent height. Typically, the registration of one altitude profile by an infrared limb sounder takes between 10s and one minute, corresponding to 80 - 500 km distance separation of adjacent profiles along the orbit track.

Infrared limb sounders have been flown on high-inclination low Earth orbits, resulting in typical orbit durations of 90-100 minutes and ∼14 orbits per day. Measurements during day and night allow for precise estimates of the zonal mean background and planetary scale waves with global wavenumbers up to ∼7 (Salby, 1982). The background atmosphere (estimated as zonal mean plus global-scale waves) is subtracted from the data and the residual temperature fluctuations are interpreted as gravity waves. Many case studies and global investigations show that GWs can be reliably inferred in this way.

For typical satellite resolution, the momentum flux can be approximated with equation (2). Applying wave analyses to the vertical profiles of temperature residuals we can infer the vertical wavelengths of the most prominent wave signatures and the corresponding temperature amplitudes. The horizontal wavelength, however, is undersampled (Ern et al., 2004). In addition, GWs in the lower stratosphere tend to occur in packets with limited horizontal extent. A good example is the mountain wave event above the Andes reported by Eckermann and Preusse (1999). Three consecutive profiles above the Andes showed distinct wave signatures with the properties of mountain waves. The profiles measured before and afterwards over the ocean do not show pronounced wave signatures. Because of this limited extent, an innovative way of determining the horizontal wavelength is required: A common dominant wave signal is identified in a pair

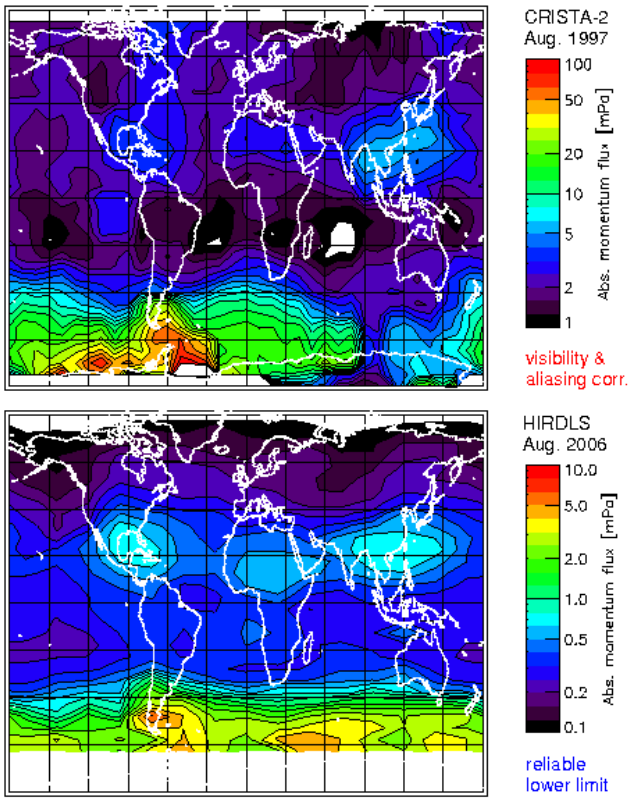

Figure 5. Global maps of gravity wave momentum flux at 20-30 km altitude derived from CRISTA (top) and HIRDLS (bottom) high resolution satellite temperature observations. CRISTA data for this map were obtained Aug. 8-16, 1997, and HIRDLS data were obtained Aug. 1-31, 2006. After Ern et al. (2004) and Alexander et al. (2008)

of adjacent vertical profiles, and horizontal wavenumber along the orbit track is inferred from the difference of the phase ϕ between the two profiles i and j with spacing $\Delta x_{i,j}$ (Ern et al., 2004):

$$k_h = \frac{\partial \phi(x_h)}{\partial x_h} = \frac{\Delta \phi_{i,j}}{\Delta x_{i,j}} \qquad (4)$$

Here k_h is the component of the total horizontal wavenumber resolved along the line joining the two profiles. Using this method global momentum flux estimates have been deduced from data obtained by two orbiting instruments: The Cryogenic Infrared Spectrometers and Telescopes for the Atmosphere (CRISTA) and the High Resolution Dynamics Limb Sounder (HIRDLS).

The CRISTA instrument was deployed on two one-week long Space Shuttle missions in November 1994 and August 1997. Vertical profiles were investigated by a combination of Maximum Entropy Method and sinusoidal fit. The momentum flux was calculated from the average vertical wavelength and temperature of the leading wave component of adjacent profiles (for details see Ern et al., 2004). The momentum flux estimates were corrected for the instrumental sensitivity and aliasing. Results are reproduced in Figure 5a. Correction factors have been applied to these "best estimate" values to approximately account for the low bias associated with unknown wave propagation directions (Ern et al., 2004). The results show very similar longitudinal variations as with momentum fluxes deduced from superpressure balloons (Hertzog et al., 2008).

The High Resolution Dynamics Limb Sounder (HIRDLS), on the EOS-Aura satellite, provided high vertical resolution temperature measurements from Jan. 2005 to Mar. 2008. Vertical profiles were analyzed by means of the S-transform and the cross-spectrum amplitude between profile-pairs was used to identify the dominant wave mode as a function of height in the profiles (Alexander et al., 2008). The horizontal wavelength was then determined using the method in (4) and shown in Figure 5b. No "best estimate" corrections were applied to the HIRDLS data so they represent lower limits of the momentum flux.

Both datasets agree well in their general structure. Largest values are found in the southern polar vortex. A second maximum is found above subtropical convection in the summer hemisphere. In the tropics, long horizontal wavelength GWs are dominant, resulting in relatively low momentum flux values. At mid and high latitudes of the summer hemisphere, weak stratospheric winds and a wind

reversal between troposphere and stratosphere strongly reduce the GW momentum flux.

In absolute values the two results are different by a factor of ∼5. CRISTA values are larger, due partly to the sensitivity/aliasing corrections applied in Ern et al. (2004), partly to the longer vertical wavelengths included in the CRISTA analysis, and as discussed in the next section, partly to intermittency and the different averaging times (1 wk vs 1 mo). By further improving the wave analysis method, calibrating the inferred values to other methods, and inferring improved corrections, new best estimates can be obtained in the future.

3.6 Quantifying wave intermittency

Gravity waves are intermittent in nature. Observations show they often occur in isolated large amplitude wave packets. For example, the August-mean map from HIRDLS observations in Fig 5b shows maximum fluxes of 6 mPa, however individual measurements exceed 100 mPa. These large amplitude events move from location to location on different days, resulting in the smaller monthly average. Such behaviour is furthermore exemplified by gravity-wave momentum-flux probability density functions (pdf). Figure 6 for instance displays momentum-flux pdfs obtained with observations collected by HIRDLS between 50S and 64S and by long-duration balloons south of 50S during Vorcore. The sporadic character of gravity-wave activity is linked to the broad tail consistently exhibited by both pdfs, which corresponds to relatively rare events with large momentum-flux values. As shown on this figure, these large values can exceed the mean momentum flux by more than one order of magnitude. The differences between the pdfs likely arise because of differences in space and time sampling (August for HIRDLS and no observation above Antarctica, September-October for Vorcore balloons with observations above Antarctica),

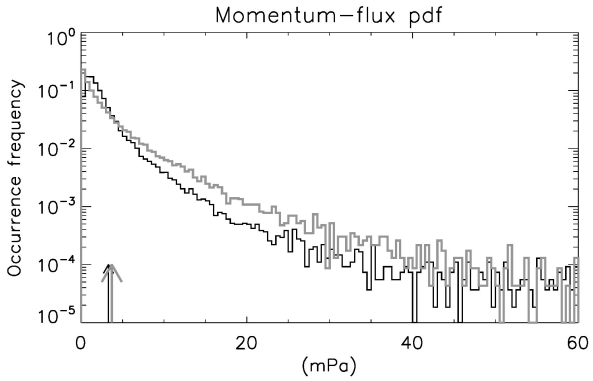

Figure 6. Gravity-wave momentum-flux probability density functions obtained with (bold gray) HIRDLS and (light black) long-duration balloon observations (see text for more details). The mean value for each pdf is indicated by the arrow at the bottom of the figure.

and also differences in detection of small amplitude gravity waves in both datasets, which impact the lowest bin of the distribution.

Average values of wave amplitude can furthermore depend on the area and time over which the average is taken. Ideally, when compared with parametrized gravity waves, observational intermittency like this would thus need to be quantified on the same grid as the global model using the parametrization. An additional problem is that waves from orographic and nonorographic sources cannot generally be clearly separated for computation of the necessary OGWD and NGWD efficiency factors.

Hertzog et al. (2008) nevertheless presented geographical information on gravity-wave intermittency in the Vorcore super-pressure balloon observations. Intermittency in this work was quantified via $\epsilon_1 = (1 + \sigma^2/\mu^2)^{-1}$ where μ is the mean momentum flux in each geographical grid box, and σ^2 the variance (cf. Figure 7). The result is labeled "probability of observation", where a value of 1 represents no intermittency, and a small value represents a high degree of intermittency.

What stands out in Fig. 7 is that the most intermittent sources, associated with the lowest probability of observation, are found above mountainous areas, such as the

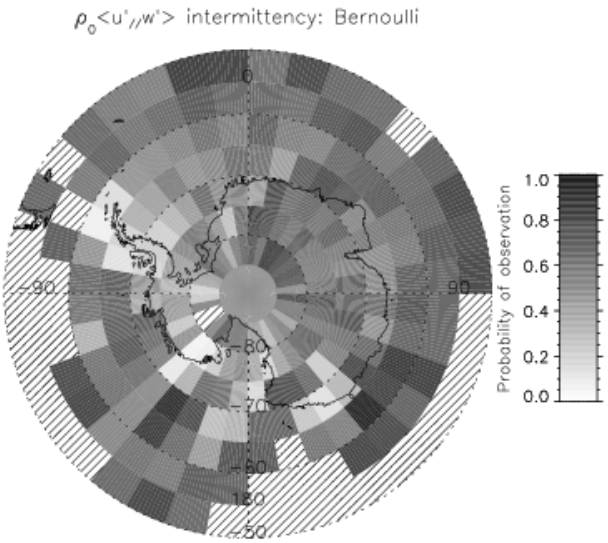

Figure 7. Wave intermittency quantified as "probability of observation" computed from super-pressure balloon data obtained during the Vorcore campaign. (Hertzog et al. (2008))

Antarctic Peninsula, the Ellsworth Range, and the Antarctic coast in several places, including Adélie Land. In contrast, the probability of observing gravity waves is largest, and the intermittency smallest, above the oceans and to a lesser extent above the Antarctic Plateau. Hertzog et al. (2008) also discussed alternate methods of defining intermittency such as ratio of the 50% percentile (the median) to the 90% percentile. The values so obtained (not shown) differ from Figure 7 in magnitude, but the patterns remain very similar.

The intermittency seen in observations arises from two causes: Sources of the waves are intermittent, and waves may be filtered as they propagate through the intervening atmosphere between their source and the observation point. There are also two types of filtering: One where the wave energy (and momentum flux) disappears because of dissipation, and the other where the wave may be refracted by changes in wind and stability to a vertical wavelength lying outside the observable range.

Gravity wave parametrizations also represent intermittency in several ways. The intermittency due to wave dissipation with height is included explicitly. For OGWD parametrizations, intermittency in sources is also included explicitly through the dependence of the wave amplitude on the local topography and the surface wind and stability. Some source intermittency is also explicit in the newer source parametrizations for wave emission from convection and fronts (e.g. Richter et al., 2010). NGWD parametrizations that are not physically linked to specific generation mechanisms lack this part of the intermittency. Both OGWD and NGWD parametrizations also include dimensionless scaling factors, such as G and ϵ (Section 2.2) that are sometimes called intermittency factors. Constraints on these parametrization factors will require careful case study comparisons.

3.7 Mapping of observations into wavenumber-frequency space

Internal atmospheric gravity waves have intrinsic frequencies between the buoyancy frequency and the Coriolis parameter, horizontal wavelengths ranging from a few kilometers to a few thousand kilometers and vertical wavelengths can theoretically vary from 0 to ∞. No single instrument can cover all these scales. However, to intercompare fluxes derived from different methods or models in order to quantify the contributions of parts of the spectrum to the whole, it is necessary to quantify the observational range of each individual instrument. If this observational filter is known, it can then be applied to model data and meaningful comparisons can be performed. Studies have shown that the observational filter not only changes the absolute value but also the relative global distributions (Alexander, 1998; Jiang et al, 2004; Ern et al. 2004; 2006; Choi et al. 2009).

It is important to keep in mind that the observational filter is not dependent on the measurement instrument alone. The first step in analyzing a dataset for GWs is

to isolate the GWs from other atmospheric signals, such as planetary scale waves. This filtering must be taken into account when calculating the observational filter. For instance, radiosondes are sensitive to almost all scales of GWs. However, GWs are determined from single profiles by a high-pass filter that rejects all vertical wavelengths longer than ∼7 km. For infrared limb sounding, on the other hand, the global background can be estimated and the visibility filter is defined by the radiative transfer and the vertical field of view of the instrument (Preusse et al., 2002).

Depending on the type of measurement, the visibility filter is best defined either in horizontal wavelength or in frequency space. Figure 8 shows both representations. The visibility filters have been converted based on the GW dispersion relation neglecting the Coriolis force, but including corrections for nonhydrostatic effects:

$$\hat{\omega}^2 = \frac{k^2 N^2}{k^2 + m^2 + 1/4H^2}, \qquad (5)$$

The conversion used typical stratospheric values of N =0.02 s^{-1} for the buoyancy frequency and H =6.5 km for the scale height.

The lines in Figure 8 give typical visibility limits. Depending on the wave amplitude and structure of the background atmosphere these sensitivity limits may shift slightly. It should also be noted that, in particular, for satellite instruments the sensitivity inside the given visibility limits is a value substantially smaller than 1. Dashed lines in Figure 8 indicate waves seen only at favorable viewing geometries. A detailed discussion of the visibility limits of various satellite instruments is given by Preusse et al. (2008). The shown visibility limits are based on the following studies: Infrared limb sounding (Preusse et al., 2002; adapted with HIRDLS minimum vertical wavelength limits), microwave sub-limb (McLandress et al., 2000; adapted for a 2.5 km vertical field of view), nadir viewing satellites (Eckermann and Wu, 2006; adapted with AIRS minimum horizontal wavelength limits), superpressure balloons (Hertzog et al., 2008), and radiosondes (Wang et al., 2005).

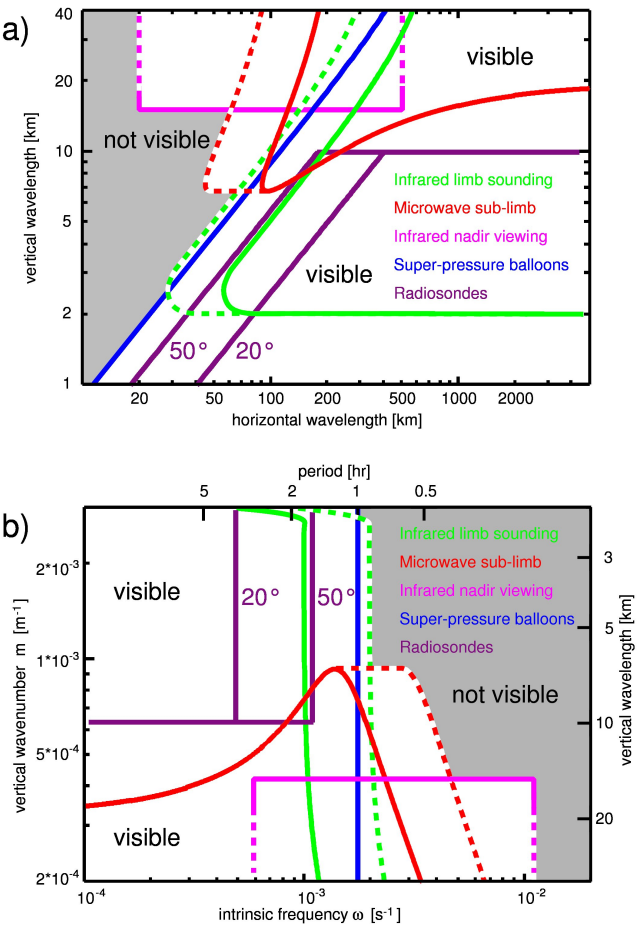

Figure 8. Typical visibility limits as functions of horizontal and vertical wavenumber (top) and frequency/vertical wavenumber (bottom) for various satellite and balloon measurement techniques. Shaded regions are not visible to any of the techniques. (After Preusse et al., 2008.)

Together, this set of measurements provides very good spectral coverage of the momentum flux over the relevant region of the spectrum as shown in Fig. 8. Gaps in coverage remain for waves with very high intrinsic frequency, although these may be partly covered by future satellite (Preusse et al., 2008) and superpressure balloon measurements. A collaborative effort among researchers with expertise in the existing measurements could lead to

a combined set of accurate constraints for parameterizations.

4 High resolution global models

Current global models used for weather prediction and climate projection must include gravity wave parameterizations because gravity waves are usually sub-grid scale phenomena or under-resolved in both the horizontal and vertical in such models. However, recent developments in supercomputing allow us to simulate gravity waves explicitly in high-resolution global climate models (e.g. Hamilton et al., 1999; Watanabe et al., 2008). Although the full spectral range of gravity waves is not completely resolved even in such high-resolution models, recent work indicates that large-scale wind and temperature fields can be realistically simulated without gravity wave parameterizations (Watanabe et al., 2008). This means that the momentum balance is internally retained between the mean fields and disturbances including resolved gravity waves, and suggests the behavior of the resolved gravity waves is also realistic in terms of propagation and dissipation in the model atmosphere. Using such high-resolution global models the effects of gravity waves can be quantitatively evaluated as functions of time and 3D space. The realism of the modelled gravity waves can be confirmed through comparison with observations by radars, lidars, radiosondes and satellites (Sato et al., 1999; Watanabe et al., 2008; Kawatani et al., 2009).

Of course, some shortcomings of these gravity-wave resolving models have to be considered. The dynamical characteristics of gravity waves depend on the model equations, e.g., hydrostatic balance is assumed. The generation of gravity waves from convection will depend on the cumulus parametrization scheme. Also, previous modelling and observational work shows that much of the momentum flux spectrum remains unresolved in such high resolution models (Hamilton, 1996; Hertzog et al., 2001; Fritts et al. 2006). In this section, some recent results from gravity-wave resolving global model simulations are reviewed.

4.1 Description of gravity-wave resolving GCMs

In order to quantify the roles of gravity waves in the global momentum budget, gravity-wave resolving (GWR)-GCMs require: 1) use of neither gravity wave parametrizations nor data assimilation of observed fields that may contaminate explicitly-resolved wave motions, 2) no extra damping of wave motions such as that typically used in numerical weather prediction models, 3) a model top near the 0.01 hPa level or higher in order to include mesospheric GWD, and 4) a spatial resolution sufficiently high to achieve realistic background wind and thermal structures, as well as realistic spatio-temporal spectra of resolved gravity waves. Indeed, GWR-GCMs satisfying one or two of these conditions have been widely used and have made significant contributions to gravity-wave statistics, generation, and propagation issues (Miyahara et al., 1986; OSullivan and Dunkerton, 1995; Sato et al., 1999; Sato, 2000; Kawatani et al., 2003, 2004, 2005; Sato and Yoshiki, 2008; Kawatani et al., 2009). To date, the Geophysical Fluid Dynamics Laboratory (GFDL) SKYHI GCM (e.g., Hamilton et al, 1999; Koshyk et al., 1999) and the JAGUAR GCM (Watanabe et al. 2008; Watanabe and Miyahara, 2009) have a model top near 0.01 hPa or higher, and likely satisfy all four requirements. According to studies with these models, a minimum resolved horizontal wavelength (λ_{hmin}) shorter than 200 km and a vertical layer thickness less than 1 km are minimal requirements in order to achieve realistic simulation of the mean winds and temperatures throughout the middle atmosphere. The GWR-GCM by Watanabe et al. (2008)

has a horizontal spectral resolution of T213 (λ_{hmin} ~188 km) and an extremely high vertical resolution (~300 m) in order to simulate realistic 3D propagation of gravity waves in sheared mean flows, as well as wave-mean flow interactions in the vicinity of critical levels. Results from this model are briefly reviewed in the rest of section 4.

4.2 Momentum budgets

Figure 9 shows meridional cross sections for the zonal mean zonal winds, E-P flux vectors and E-P flux divergence in July, which are simulated in the GWR-GCM by Watanabe et al. (2008). The meridional structures of the southern hemisphere polar vortex and the northern hemisphere summertime easterlies in the stratosphere and mesosphere are well simulated. The wave forcing due to simulated gravity waves with horizontal wavelengths of 188–950 km (GW group) dominates the total wave forcing in the mesosphere of both hemispheres, which strongly decelerates the upper part of the polar vortex and the summertime easterlies to maintain their characteristic meridional structures. The meridional distribution of the E-P flux in Fig. 9 indicates that westward propagating gravity waves originating near 40°–50°S of the troposphere propagate upward and slightly southward to reach near 50°–60°S in the mesosphere. The characteristic propagation results in the westward forcing having a peak near 50°–60°S that maintains the characteristic double-peak structure of the mesospheric westerlies (Sato et al., 2009a; Watanabe et al., 2009). Similarly, eastward propagating gravity waves originating near 10°–20°N of the troposphere propagate upward and slightly northward to reach ~ 15°–25°N of the mesosphere, causing a eastward forcing maximum at these latitudes. The eastward forcing with characteristic structure is likely responsible for the summertime mesospheric easterlies tilting toward higher latitudes (Sato et al., 2009a). It should be noted that gravity wave forcing is also dominant in the summer stratosphere. Watanabe (2008) compared the meridional distributions of zonal momentum flux and zonal gravity wave forcing in the GWR-GCM simulation to those calculated using the Hines gravity wave parametrization scheme. They showed that neglect of the lateral propagation of gravity waves, including the effects of lateral spreading of the gravity wave momentum fluxes from the source locations in the gravity wave drag parametrization, could cause significant biases in the parametrized gravity wave forcing.

4.3 Seasonal variation of gravity waves and their sources

An advantage of global model studies is the ability to identify gravity wave sources in the model (Sato et al., 2009a). The sources of gravity waves that act to maintain the weak wind layer in the upper mesosphere have not been known up until now.

It is known from radar observations in middle latitudes that gravity wave activity in the mesosphere exhibits a semiannual variation having two maxima at solstices, while that in the lower stratosphere shows an annual variation with a winter maximum (Vincent and Reid 1983: Tsuda et al., 1990; Sato 1994). Observations also indicate that the momentum flux $\bar{\rho}\overline{u'w'}$ in the mesosphere is positive in summer and negative in winter. A likely explanation for the seasonal cycle is critical level filtering of gravity waves in the seasonally varying background winds (e.g., Fukao, 2007). The seasonal cycle of momentum fluxes simulated by the high-resolution model (GWR-GCM, Watanabe et al., 2008) is shown as a function of latitude in the mesosphere and lower stratosphere in Figure 10 (Sato et al., 2009a). It is seen that $\bar{\rho}\overline{u'w'}$ exhibits an annual variation which is negative in winter and positive in summer both in the lower stratosphere and the mesosphere, and the seasonal cycle in mid-latitudes is consistent with radar observations. In the lower stratosphere,

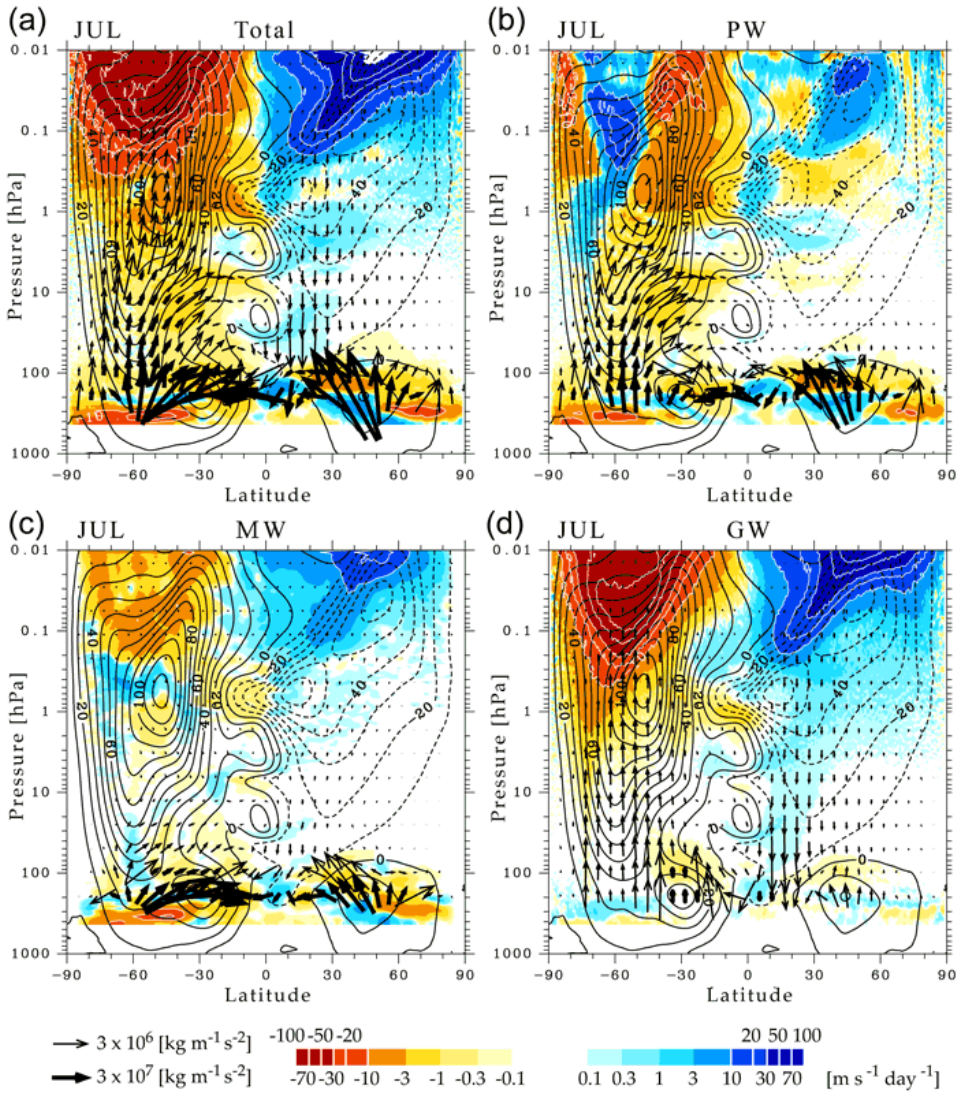

Figure 9. E-P flux vectors (arrows) and eastward accelerations of zonal mean zonal wind due to divergence of E-P flux (colors) for July (average). (a) Total wave components, (b) PW group with zonal wavenumber (k) =1-3, (c) MW group with horizontal wavelengths greater than 950 km excluding the k = 1-3 components, (d) GW group with horizontal wavelengths of 188-930 km. Vertical component of E-P flux is multiplied by 250. Scales of arrows are modified for clarity, depending on magnitude of E-P flux. Color scale is logarithmic. Contours denote zonal mean zonal wind in 10 m s^{-1} intervals (Watanabe et al., 2008).

flux maxima move from regions with high topography in winter to monsoon convective regions in summer, suggesting different sources in winter and in summer. An interesting feature is that the latitudes of the momentum flux maxima at the respective solstices are higher in the mesosphere than in the lower stratosphere. This is partly due to gravity wave filtering by the mean wind, but also due to gravity wave propagation toward the jet core, as shown in the next section. Thus, it is found from the model results that the annual variation of gravity wave fluxes in the mesosphere is caused by both seasonal source variations and by critical level filtering.

The horizontal distribution of gravity wave sources is examined using a horizontal map of $\bar{\rho}\overline{u'w'}$. Results for July are shown in Fig. 11. In the winter (southern) hemisphere, there are several regions with largely negative $\bar{\rho}\overline{u'w'}$ around the high mountains such as the Andes, Antarctic Peninsula, and the east coast of Australia. In addition to these isolated regions, negative $\bar{\rho}\overline{u'w'}$ is distributed zonally along the eastward jet. These may be

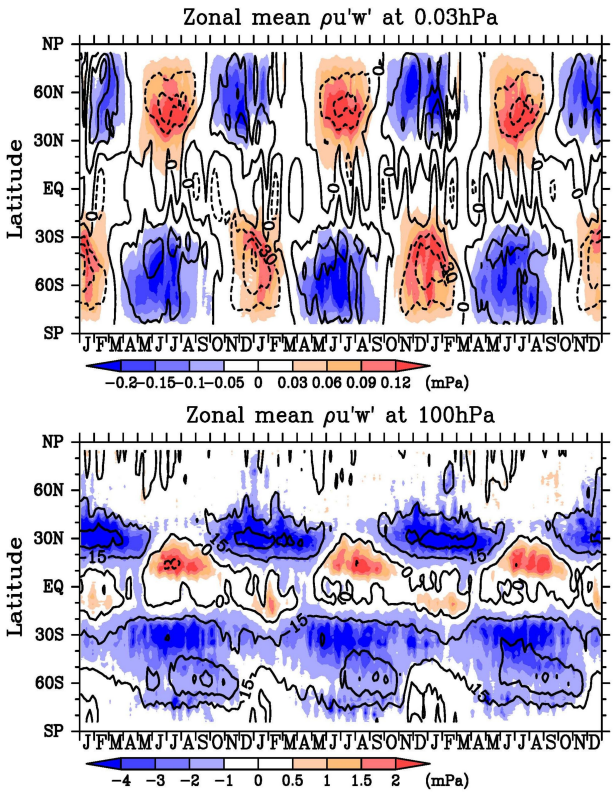

Figure 10. Time-latitude cross-sections of momentum flux associated with gravity waves (colours) and longitudinal mean eastward winds (contours) at 0.03 hPa in the upper mesosphere and at 100 hPa in the lower stratosphere (Sato et al., 2009a)

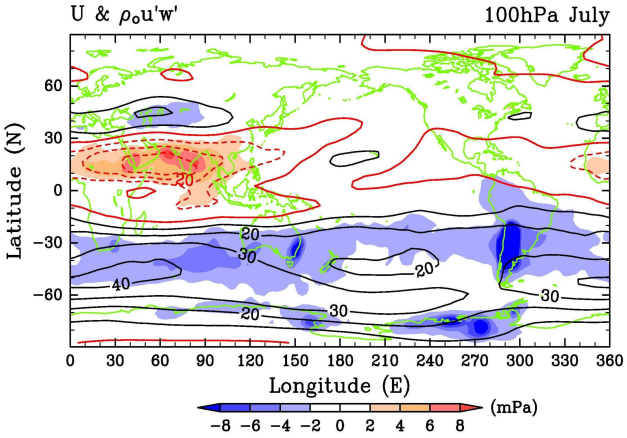

Figure 11. A horizontal map of momentum flux associated with gravity waves (colours) at 100 hPa in the lower stratosphere. Red solid, red dashed and black contours indicate zero, westward and eastward winds in m s^{-1} (Sato et al., 2009a).

due to the gravity waves excited by spontaneous adjustment around jet-frontal systems (OSullivan and Dunkerton, 1995; Kawatani et al. 2004; Zhang, 2004; Plougonven and Snyder. 2007; Sugimoto et al. 2008; Sato and Yoshiki, 2008; Tateno and Sato, 2008). In the summer (Northern) hemisphere, strong positive $\bar{\rho}\overline{u'w'}$ is observed only in the Indian and African monsoon regions. Westward winds are dominant in the lower stratosphere of this region. Gravity waves having positive $\bar{\rho}\overline{u'w'}$ generated by vigorous convection generally penetrate into the middle atmosphere only through this westward wind region, an open window to the middle atmosphere. In January, a similar dominance of positive $\bar{\rho}\overline{u'w'}$ is observed in the Australian monsoon region (not shown). These model results suggest that gravity waves sources are not distributed uniformly in the horizontal and have significant seasonal variations, unlike the assumptions used in most gravity wave parametrizations. If these model variations in sources are realistic, they may be important in explaining the asymmetric structure of the mesospheric jets in winter and summer, as described in Section 4.4.

4.4 Three dimensional propagation of gravity waves

As illustrated in most standard text books, topographically-forced gravity waves in a steady wind regime have zero horizontal group velocity. OGWD parametrizations assume this explicitly, however zero horizontal group velocity is true only in a two dimensional hydrostatic framework. In three dimensional space, topographically-forced gravity waves propagate leeward (e.g. Shutts, 1995; Preusse et al. 2002). This is because the energy of topographically-forced gravity waves can be freely advected by the background wind component perpendicular to the wavenumber vector, although the intrinsic group velocity is balanced with the background wind component parallel to the wavenumber vector (Sato et al. 2009b). The model simulation and a ray-tracing analysis indicate that with this mechanism, gravity waves from the Andes propagate leeward over a longitudinal

distance of about 70°. Another important aspect of lateral propagation is the focusing of gravity waves into the jet. The modification of the wavenumber vector by the latitudinal gradient of the mean wind described by Dunkerton (1984) and Preusse et al. (2002) acts on gravity waves with negative $\bar{\rho}\overline{u'w'}$ in the eastward jet. Similar wave focusing occurs for gravity waves with positive $\bar{\rho}\overline{u'w'}$ in westward jets (Eckermann, 1992; Sato et al., 2009a).

4.5 Gravity waves as a driving force for the QBO

Atmospheric general circulation models (AGCMs) are effective tools with which to study the roles of atmospheric waves in driving the QBO (Takahashi 1996, 1999; Horinouchi and Yoden 1998; Hamilton et al., 1999, 2001; Giorgetta et al., 2002, 2006; Shibata and Deushi 2005; Kawatani et al., 2005, 2010a, 2010b). Kawatani et al. (2010a) examined the relative contribution of equatorially trapped wave modes (EQWs) and internal gravity waves in driving the QBO using outputs over 3 model years from the T213L256 GWR-GCM (Watanabe et al., 2008). Here EQWs are defined with order $n = -1$ to 2 with zonal wavenumber $s \leq 11$ and equivalent depths of 2–90m, while gravity waves have $s \geq 12$. (Kawatani et al., 2010a) Figure 12a presents a time-height cross section of the EP-flux divergence due to all resolved wave components averaged for 10°S–10°N. A QBO-like oscillation having a period of approximately 15 months is obvious. The period is shorter than the observed value, probably partly due to the model's underestimation of mean ascent in the tropics. The amplitude and bottom levels of the QBO-like oscillation are realistic, however. It is also clear that eastward wave forcing corresponds well to the eastward shear of the QBO, while westward forcing corresponds to westward shear, indicating that the waves in the model certainly drive the QBO-like oscillation. It is important to note that the relationship between wave forcing and vertical zonal wind shear is also clear in the SAO around the stratopause.

Figure 12b show the time series of the EP-flux divergence due to all waves, eastward EQWs, westward EQWs, and gravity waves as well as the forcing due to the residual circulation at 30 hPa averaged for 10°S–10°N. Generally, the forcing due to the residual circulation will be opposite to, and smaller than, the total wave forcing. In the eastward shear phase, peaks of eastward forcing due to the eastward EQWs and gravity waves are almost coincident. The contributions of eastward EQWs and gravity waves to the eastward forcing are ∼25–50% and ∼50–75%, respectively. In the westward shear phase, westward EQWs contribute ∼10% at most to QBO driving. On the other hand, the contribution by Rossby waves propagating from the winter hemisphere was about 10–20% (not shown). Consequently, gravity waves play a crucial role in driving the QBO in its westward shear phase.

In order to investigate the horizontal scales of gravity waves driving the QBO, time series of EP-flux divergence are shown separately for the components of $12 \leq s \leq 42$, $43 \leq s \leq 106$, and $107 \leq s \leq 213$ (Fig. 12c). The eastward forcing due to each component is comparable in the eastward shear phase. In contrast, components with $42 \leq s \leq 213$ (horizontal wavelength ≤∼1000 km) are dominant for westward forcing in the QBO westward shear phase. These results are in general agreement with earlier studies using parametrized gravity waves (Giorgetta et al., 2002), but provide far more detail.

Recent satellite and modelling studies indicate that wave activity depends greatly on both latitude and longitude in the equatorial region (Alexander et al., 2008; Ern et al., 2008; Kawatani et al., 2009). Using outputs from

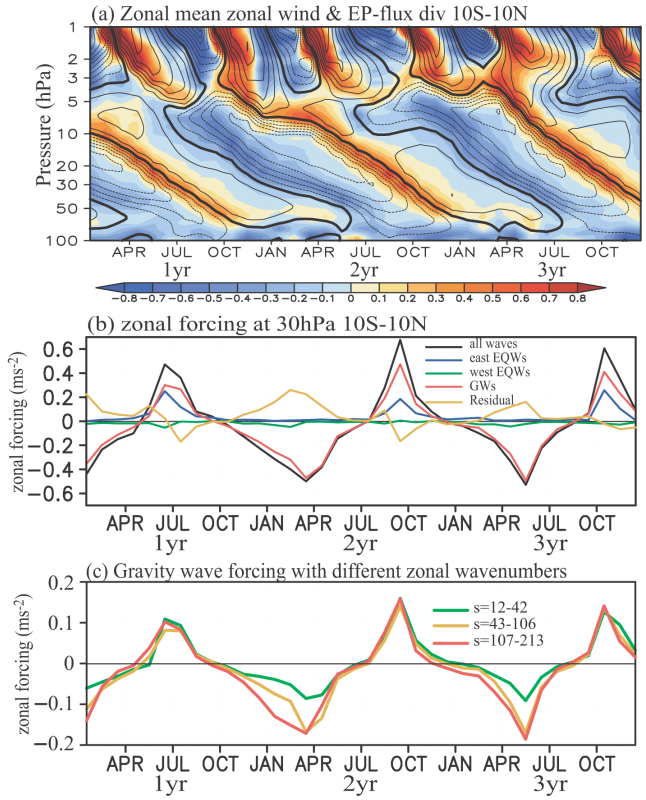

Figure 12. Time-height cross section of zonal-mean zonal wind (contour) and EP-flux divergence (shaded) at 10°S10°N. Red and blue colors correspond to eastward and westward forcing, respectively. The contour and shading intervals are 5 m s^{-1} and 0.1 m s^{-1} day^{-1}, respectively. Eastward and westward winds are shown with solid and dashed lines. (b) Time variation of EP-flux divergence due to all waves (black), eastward EQWs (blue), westward EQWs (green), gravity waves (red), and forcing due to the residual circulation (yellow) at 30 hPa averaged from 10°S to 10°S. (c) The same as (b), but for EP-flux divergence due to $12 \leq s \leq 42$ (green), $43 \leq s \leq 106$ (yellow), and $107 \leq s \leq 213$ (red). Note that the range of the ordinate axis of (c) is different from that of (b). (After Kawatani et al. (2010a))

the T213L256 AGCM Kawatani et al. (2010b) focused on the 3D distribution of wave forcing. In both eastward and westward shear phases of the QBO, EP-flux divergence due to gravity waves is due mainly to the vertical component of EP-flux, which suggests that a gravity wave drag parametrization assuming only vertical wave propagation is appropriate, at least for a QBO simulation. However, the divergence of meridional components of EP-flux associated with EQWs is comparable to the contribution of vertical components where vertical shear in QBO zonal winds is large. The longitudinal dependence of wave forcing was examined using 3D wave activity flux applicable to inertia-gravity waves derived by Miyahara (2006). It is shown that the Walker circulation plays a crucial role in filtering waves propagating into the lower stratosphere – westward winds in the Eastern Hemisphere of the upper troposphere prevent westward waves from entering the stratosphere, and eastward winds in the Western Hemisphere block eastward waves – a generalization of the Kelvin wave result reported by Ryu et al. (2008). In the eastward shear zone of the QBO, the eastward forcing due to gravity waves in the Eastern Hemisphere is much larger than that in the Western Hemisphere. On the other hand, in the westward wind shear zone, westward wave forcing does not vary much in the zonal direction, despite slightly larger westward wave forcing in the Western Hemisphere than in the Eastern Hemisphere. Kawatani et al. (2010b) concluded that the zonal variation of wave forcing in the stratosphere results from three factors: (1) zonal variation of wave sources, (2) vertically sheared zonal winds associated with the Walker circulation, and (3) the phase of the QBO.

5 Data assimilation techniques for estimating the Gravity wave momentum budget

In section 3, high resolution observations were used to determine the momentum flux transported by small-scale waves. An alternative approach is to estimate the missing force in low-resolution analyses using time-averaged observations and zonal-mean balance conditions (e.g. Shine (1989), Marks (1989)) and resolved wave forcing (Alexander and Rosenlof, 1996). More recently, data assimilation methods have been used to capture the 3D, time evolving drag field. The actual missing force due to unresolved waves is the result of interactions between the mean flow, the resolved waves and the unresolved waves. In particular, the forcing by unresolved waves changes the

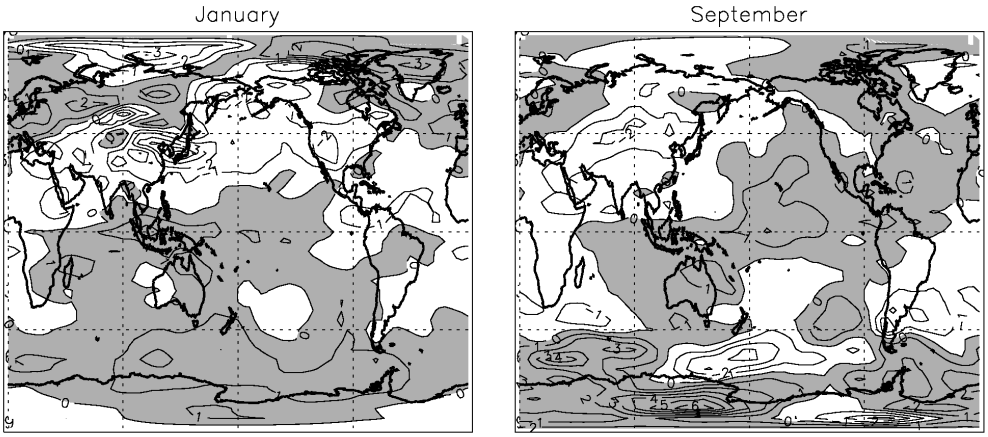

Figure 13. Zonal momentum flux (contour interval is 0.01 Pa, positive zonal flux is shaded) at 100 hPa for January and September 2002. After Pulido and Thuburn (2008).

mean flow which then produces changes in the forcing by the resolved waves. By constraining the large scale flow with measurements (using data assimilation), the forcing due to resolved waves is known so that the misfit between the observed and forecasted drag can be attributed to unresolved waves. Thus, data assimilation methods have the unique advantage of being able to separate the response of the drag due to resolved waves from that due to unresolved waves. In this section, the standard assimilation problem is introduced before considering its application to the inverse problem of estimating the missing force due to unresolved gravity waves.

5.1 The data assimilation problem

Data assimilation combines measurements and model forecasts according to their accuracies and produces a "best" estimate of the state of the atmosphere on a model grid. This definition is appropriate to the process performed at numerical weather prediction centres. Mathematically, the process can be represented as follows. The assimilation step seeks a model state, $\mathbf{x}$, which best fits observations, $\mathbf{z}$, and which is not too far from a background model state, $\mathbf{x_b}$. This can be done by minimizing the following cost function:

$$J(\mathbf{x}) = \frac{1}{2}(\mathbf{x}-\mathbf{x}_b)^T\mathbf{B}^{-1}(\mathbf{x}-\mathbf{x}_b) + \frac{1}{2}(\mathbf{z}-H(\mathbf{x}))^T\mathbf{R}^{-1}(\mathbf{z}-H(\mathbf{x})). \quad (6)$$

Here H is a possibly nonlinear operator mapping model to observed variables, and $\mathbf{B}$ and $\mathbf{R}$ are error covariance matrices corresponding to the model and observed state vectors. The background model state, $\mathbf{x}_b$, typically comes from a short term (6 or 12-h) forecast valid at the time of the analysis. If the accuracy of measurements and background state is known, then the state which minimizes (6) is optimal in that it minimizes the analysis error variance. However, error covariances (particularly those associated with the background state) can never really be known and must, at best, be modelled approximately. Thus the optimality principle holds only in theory. Once an initial state is obtained, a short term model forecast can be launched to obtain a background state for a later assimilation. Then the assimilation cycles continuously every 6 or 12 hours through alternating assimilation and forecast steps. (6) rather precisely defines the so-called 3D variational assimilation procedure (or 3D-Var). By allowing H to include temporal interpolation of the model state to the time of observations distributed over a time window

using the forecast model, this equation can also describe the so-called 4D variational assimilation procedure (or 4D-Var). Finally, note that Optimal Interpolation and 3D-Var are formally equivalent in the case of linear observation operators and Gaussian observation and background errors. Thus (6) can also describe the Optimal Interpolation scheme.

5.2 Evidence that the mesosphere is constrained by measurements from the troposphere and stratosphere

Gravity wave drag reaches large amplitude in the upper mesosphere and lower thermosphere region. Since observations that are assimilated into weather and climate models are primarily from the troposphere and stratosphere, it is important to know whether such observations can constrain the mesosphere. Conversely, if the mesosphere is slaved to the lower atmosphere then the misfit of measurements and forecasts in the mesosphere can be used to retrieve information about the troposphere and stratosphere. Recently, Ren et al. (2008) have shown that forecasts of the 2002 southern hemisphere stratospheric sudden warming that capture the vortex splitting event also produce a mesospheric cooling while forecasts that do not produce a vortex splitting have much weaker mesospheric cooling. In these simulations, observations were used only below 1 hPa so that model forecasts in the mesosphere were due to the model response to data insertion. Of the mechanisms available to produce a mesospheric cooling during forecasts, parametrized GWD was shown to be the dominant one. Thus, in a dramatic event such as a sudden warming, GWD is playing an important role in determining the mesospheric forecast. While this results demonstrates that the mesosphere is slaved to the lower atmosphere in at least one warming, a more statistical result was obtained by Nezlin et al. (2009). In this work, observations from the troposphere and stratosphere were able to define the large scales (below wavenumber 10) in the mesosphere. Since observations were simulated from a model run, there was an underlying assumption that the model is realistic. Even with this assumption, and with perfect observations (no random errors added), mesospheric scales higher than wavenumber 10 could not be determined. Thus only the large scale mesospheric flow can be constrained by observations from below the mesosphere.

The role of initial conditions in the troposphere and stratosphere provided by data assimilation in controlling a subsequent mesospheric forecast without a local data assimilation constraint was similarly demonstrated in studies by Coy et al. (2005) and Siskind et al. (2007). These earlier studies also validated their mesospheric forecasts against independent SABER temperature observations. The observational constraints showed that the mesospheric response to stratospheric warmings can be quite variable from event to event, but the response is typically characterized by a shallow lower mesospheric cooling layer that then transitions to an upper mesospheric or lower thermospheric secondary warm layer response. These details differ from the Ren et al. (2008) study, and they may be sensitive to the details of the GWD parameterization used in the different models.

5.3 Using data assimilation to estimate gravity wave drag morphology

The response of the large-scale extratropical flow to an external localized forcing is non-local and may include geostrophic and transient components (see Fig. 6 in Haynes et al. 1991, also Fig. 1 in Pulido and Thuburn 2005). If the initial state is known, then an initial uncertain forcing added to the momentum equations will lead to a departure of the predicted drag (or wind) from the

observed drag (or wind). Inverse techniques use this mismatch of the observed and predicted drag and knowledge of the flow evolution (model equations) to trace back and recover the initial localized forcing. Pulido and Thuburn (2005) were the first to apply an inverse technique based on the concepts of 4D variational data assimilation to estimate GWD. A cost function that measures the difference between the state of the dynamical model and observations was defined:

$$J(\mathbf{x}_0, \mathbf{X}) = \frac{1}{2} \sum_i (\mathbf{z}_i - M_i(\mathbf{x}_0, \mathbf{X}))^T \mathbf{R}^{-1} (\mathbf{z}_i - M_i(\mathbf{x}_0, \mathbf{X})) \tag{7}$$

where $\mathbf{z}_i$ are the observations (the analysis in this context) at time i, $\mathbf{x}_0$ is the initial condition, $\mathbf{X}$ is the GWD, and M is the dynamical model so that $M_i(\mathbf{x}_0, \mathbf{X})$ gives the state of the model at time i. H does not appear explicitly in (7) because observations $\mathbf{z}_i$ were first transformed to model state variables. The assumption of complete ignorance of the gravity wave drag is taken so that there is no background term (first term in (7)) in the cost function. As the observations and the initial conditions $\mathbf{x}_0$ are known, the only unknown field in the cost function (7) is the GWD $\mathbf{X}$. The determination of the minimum of the cost function gives the optimum gravity wave drag.

Budget techniques give a gravity wave deceleration centre in the winter hemisphere and a weaker deceleration centre in the summer hemisphere above the stratopause, while gravity waves accelerate the jets in the stratosphere (e.g. Alexander and Rosenlof, 1996). With data assimilation, the estimated zonal mean of the zonal component of GWD qualitatively resembles the results obtained by budget studies. However, there are quantitative differences such as stronger and more concentrated deceleration centres being found with the data assimilation technique (Pulido and Thuburn 2006). While the source of these differences remains unclear, one advantage of assimilation techniques over budget studies is clear: that they are able to infer 3D, day-to-day GWD fields, of both zonal and meridional components.

Using the GWD estimated with data assimilation, Pulido and Thuburn (2008) inferred the momentum flux at 100 hPa assuming that momentum fluxes are negligible above 0.24 hPa. The calculated momentum flux is dominated by the GWD in the lower stratosphere where model errors may have some impact on the GWD estimation. Figure 13 shows the monthly mean momentum flux at 100 hPa for September 2002. Maximum momentum fluxes are found over Antarctica ($\sim +50$ mPa) and over South America (~ -20 mPa) which could be related to the orography or the storm track location. During boreal winter, maximum momentum fluxes are located over Asia (~ -40 mPa).

The assimilation-estimated GWD also has reasonable features at low latitudes where GWD is expected to play a role in the quasi-biennial oscillation (QBO). With a one-year GWD estimate, Pulido and Thuburn (2008) noted that at low latitudes the forcing changes from positive to negative at 40 hPa (both are accelerating the magnitude of zonal mean flow, see Fig. 14) in consonance with the zonal wind that is changing from the positive to the negative phase of the QBO. At higher altitudes a semiannual oscillation is found in the zonal GWD, however the GWD has opposite phase to the zonal wind semiannual oscillation, and is superimposed on a mean positive forcing (left and middle panels of Fig. 14).

Estimation of GWD with data assimilation relies on two major assumptions: The forcing produced by gravity waves must be larger than model error and it must be a large-scale and systematic forcing so that low-resolution

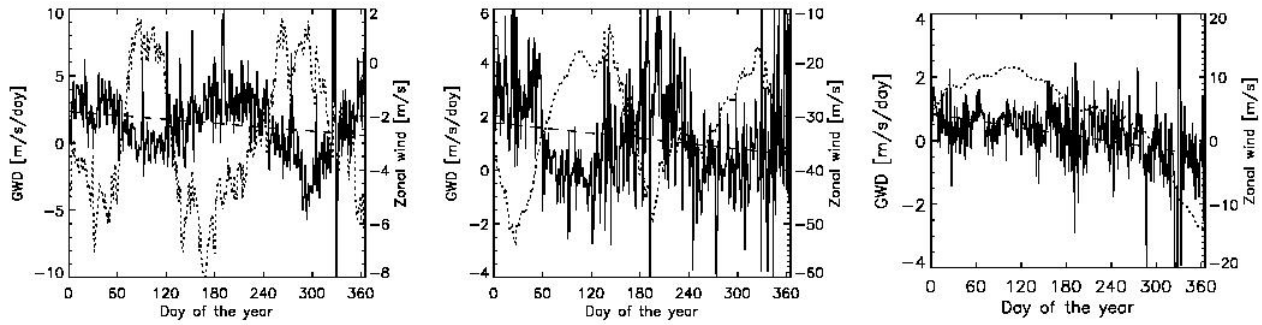

Figure 14. Daily zonal mean GWD (continuous line) estimated at the Equator with a data assimilation technique for 2002, zonal mean wind (dotted line) and linear regression of the drag (dashed line) at 0.24 hPa (left), 1.15 hPa (middle) and 40 hPa (right panel). After Pulido and Thuburn, 2008.

observations can capture the response of the flow to the forcing. The latter assumption has a physical foundation: the systematic filtering of gravity waves by background winds in the lower atmosphere produces a large scale, slow forcing in the mesosphere. The former assumption is justifiable since climate models without a gravity wave parametrization or Rayleigh friction are expected to contain a strong bias above the stratopause, the so-called missing force (e.g. Holton 1982). Since a poor representation of other physical processes in models (e.g. radiation) will also be attributed to GWD, it is important that GWD be the dominant source of model error. The phasing of these annual and semiannual variations are in agreement with monthly means derived in Alexander and Rosenlof (2003).

Thus far, assimilation techniques have been applied to estimate the drag due to unresolved gravity waves. In the future, these techniques may also be applied to directly determine parameters used in GWD schemes. The use of data assimilation or inverse techniques offers an objective way to estimate climate model parameters. Such an objective parameter estimation may be of paramount importance to realistically represent the QBO, a circulation feature that is still not well represented in state-of-the-art climate models.

6 Summary

Here we summarize a few key points regarding the effects of gravity waves in climate models, results from observational analyses that constrain these effects, and very high-resolution model studies that highlight both the capabilities of these models as well as the challenges that still lie ahead for realistic treatment of gravity wave effects on the large-scale circulation.

Parametrized orographic gravity wave drag (OGWD) has strong effects on mean winds and planetary wave propagation in the winter hemisphere stratosphere, and it is through this mechanism that OGWD affects climate change response patterns in models. OGWD also has a strong effect on variability both in the stratosphere and at the surface. An important tuning parameter controlling these effects (ϵ), which scales the drag profile, remains difficult to quantify with observations. Angular momentum conservation is a critical constraint that must be applied to ensure robust results in climate change studies that use parametrized GWD.

Recent studies have noted parametrized GWD effects on some climate responses in models. OGWD affects predicted trends in the Brewer-Dobson circulation in climate change scenarios. Changes in this circulation have repercussions for the transport of trace constituents and

for ozone chemistry. Parametrized non-orographic gravity wave drag (NGWD) plays a crucial role in chemistry-climate forecasts of ozone recovery where it is used to control polar temperatures in model studies, particularly in the southern hemisphere. Small changes to the tuning parameter that sets the source-level momentum flux can have very large effects on the seasonal development of the ozone hole. Source parametrizations for known sources such as convection, fronts, and jet imbalance are needed to give gravity waves sensitivity to meteorology in climate models so they can evolve with changing climate. Although experimental versions of such parametrizations exist, they are currently poorly constrained.

Global-scale distributions of gravity wave momentum flux and other propagation properties have been derived from both satellite- and balloon-borne measurements. Each measurement and analysis method observes only a portion of the full spectrum of momentum fluxes, but the combined set of measurements can provide very good spectral coverage. Momentum fluxes derived from observations can also depend on the resolution at which they are reported because wave events can be very localized in both space and time, a concept termed intermittency. Because of intermittency, local values can be more than an order of magnitude larger than areal or time-averaged fluxes.

Time-averaged global observations of momentum fluxes derived from different measurements are showing some similarities: Largest fluxes in the winter hemisphere over topography, enhanced fluxes at low latitudes and in the summer hemisphere over known centers of deep convection. Uncertainties in magnitude and direction of propagation remain, however methods for intercomparison and correction for various limitations of the different observations are being developed that should allow a more quantitative global picture to emerge in the very near future. Future measurements with improved resolution that extend the global coverage of the balloon data and that quantify the wave propagation directions in satellite data will have major impacts on reducing uncertainties.

Satellite evidence suggests OGWD may play a larger role in the southern hemisphere than previously thought because of small island sources of mountain waves. The Southern Andes and Antarctic Peninsula also appear in all the observations and models as a locus of maximum fluxes. Nonhydrostatic propagation effects and nonlinear processes, both of which are neglected in current OGWD parametrizations, likely play important roles here. Future work may lead to better methods for treatment of these processes in the parametrizations.

The most recent gravity-wave resolving high-resolution global climate models have achieved a high level of realism, and the gravity wave momentum fluxes in these models show excellent agreement with observations from radar and satellite. This is somewhat surprising because observational evidence suggests that a large fraction of the gravity wave momentum flux remains unresolved in these models. The 3D picture afforded by the model allows calculation of global distributions of the vector momentum flux, the associated GWD, and the sources of the gravity waves. The high-resolution model results suggest that waves in the gravity wave part of the spectrum provide more than half of the wave forcing necessary to drive the QBO.

Patterns in the gravity wave momentum flux derived from data assimilation methods show some consistency with the observations and similarity to patterns derived from high resolution models. These similarities are

encouraging since the methods are all dramatically different. The similarities among the global distributions derived from these very different methods suggest that a quantitative picture of the sources of gravity waves and the global distributions of momentum flux will emerge in the near future.

Tightening constraints on a GWD parametrization will require the additional step of translating these findings into practical limits on the set of tuning parameters, which differs from one parameterization to another. This can only be accomplished through collaborative research between model developers and observational scientists.

Acknowledgement

This work is part of an ongoing activity within the World Climate Research Programme's (WCRP) Stratospheric Processes and their Role in Climate (SPARC) project. The authors thank SPARC International Project Office for its help with organizing and facilitating the workshop where this manuscript was initiated. Additional support for MJA came from NASA's Earth Science Mission Directorate (contract #NNH08CD37C) and the National Science Foundation's Physical and Dynamic Meteorology Program (Award #0632378), and for SDE and MJA from NASA contract #NNH08AE43I.

References

Alexander, M. J. and K. H. Rosenlof, 1996: Nonstationary gravity wave forcing of the stratospheric zonal mean wind. *J. Geophys. Res.*, **101**, 23465-23474.

Alexander, M. J., 1998: Interpretations of observed climatological patterns in stratospheric gravity wave variance. *J. Geophys. Res.*, **103**, 8627-8640.

Alexander, M. J. and K. H. Rosenlof, 2003: Gravity wave forcing in the stratosphere: Observational constraints from UARS and implications for parameterization in global models, *J. Geophys. Res.*, **108**, D19 ,doi:10.1029/2003JD003373.

Alexander, M. J., and C. Barnet, 2007: Using satellite observations to constrain parameterizations of gravity wave effects for global models, *J. Atmos. Sci.*, **64**, 1652–1665.

Alexander, M. J. and H. Teitelbaum, 2007: Observation and analysis of a large amplitude mountain wave event over the Antarctic Peninsula, *J. Geophys. Res.*, **112**, D21103, doi:10.1029/2006JD008368.

Alexander, M. J., S. D. Eckermann, D. Broutman and J. Ma, 2009: South Georgia island wave patterns observed in the stratosphere via satellite, *Geophys. Res. Lett.*, **36**, L12816, doi:10.1029/2009GL038587.

Alexander S. P., T. Tsuda, Y. Kawatani, and M. Takahashi, 2008: Global distribution of atmospheric waves in the equatorial upper troposphere and lower stratosphere: COSMIC observations of wave mean flow interactions. *J. Geophys. Res.*, **113**, D24115. doi:10.1029/2008JD010039.

Allen, S. J., and R. A. Vincent, 1995: Gravity wave activity in the lower atmosphere: Seasonal and latitudinal variations. *J. Geophys. Res.*, **100**, 1327-1350.

Boccara, G., A. Hertzog, R. A. Vincent, and F. Vial, 2008: Estimation of gravity-wave momentum fluxes and phase speeds from quasi-Lagrangian stratospheric balloon flights. 1: Theory and simulations, *J. Atmos. Sci.*, **65**, 3042-3055.

Brewer, A. M., 1949: Evidence for a world circulation provided by measurements of helium and water vapor distribution in the straosphere, *Q. J. R. Meteorol. Soc.*, **75**, 351-363.

Butchart, N., A. A. Scaife, M. Bourqui, J. de Grandpré, S. H. E. Hare, J. Kettleborough, U. Langematz, E. Manzini, F. Sassi, K. Shibata, D. Shindell, and M. Sigmond, 2006: Simulations of anthropogenic change in the strength of the Brewer-Dobson circulation. *Climate Dyn.*, **27**, 727–741, doi:10.1007/s00382-006-0162-4.

Butchart, N., I. Cionni, V. Eyring, D. W. Waugh, H. Akiyoshi, J. Austin, C. Brühl, M. P. Chipperfield, E. Cordero, M. Dameris, R. Deckert, S. M. Frith, R. R. Garcia, A. Gettelman, M. A. Giorgetta, D. E. Kinnison, F. Li, E. Mancini, C. McLandress, S. Pawson, G. Pitari, D. A. Plummer, E. Rozanov, F. Sassi, J. F. Scinocca, T. G. Shepherd, K. Shibata, and W. Tian, 2010: Chemistry-climate model simulations of 21st century stratospheric climate and circulation changes, *J. Clim.* (in press).

Chandran, A., D. W. Rusch, S. E. Palo, G. E. Thomas, and M. J. Taylor, 2009: Gravity wave observations in the summertime polar mesosphere from the Cloud Imaging and Particle Size (CIPS) experiment on the AIM spacecraft, J. Atmos. Sol.-Terr. Phys., 71, 392–400.

Chen, G. and P. Zurita-Gotor, 2008: The tropospheric jet response to prescribed zonal forcing in an idealized atmospheric model. *J. Atmos. Sci.*, **65**, 2254–2271.

Choi, H.-J., H.-Y. Chun, and I.-S. Song, 2009: Gravity wave temperature variance calculated using the ray-based spectral parameterization of convective gravity waves and its comparison with Microwave Limb Sounder observations. *J. Geophys. Res.*, **114**, D08111, doi:10.1029/2008JD011330.

Chun, H.-Y., I.-S. Song, J.-J. Baik, and Y.-J. Kim, 2004: Impact of a convectively forced gravity wave drag parameterization in NCAR CCM2. *J. Clim.*, **17**, 3530–3547.

Coy, L. D. E. Siskind, S. D. Eckermann, J. P. McCormack, D. R. Allen, and T. F. Hogan, 2005: Modeling the August 2002 minor warming event. *Geophys. Res. Lett.*, **32**, L07808, doi:10.1029/2005GL022400.

Dewan, E.M., R. H. Picard, R. R. O'Neil, H. A. Gardiner, J. Gibson, J. D. Mill, E. Richards, M. Kendra, and W. O. Gallery, 1998: MSX satellite observations of thunderstorm-generated gravity waves in mid-wave infrared images of the upper stratosphere. *Geophys. Res. Lett.*, **25**, 939-942.

Dunkerton, T. J., 1984: Inertia-gravity waves in the stratosphere. *J. Atmos. Sci.*, **41**, 3396-3404.

Eckermann, S. D., 1992: Ray-tracing simulation of the global propagation of inertia gravity waves through the zonally averaged middle atmosphere. *J. Geophys. Res.*, **97**, 15,849-15,866.

Eckermann, S. D. and P. Preusse, 1999: Global measurements of stratospheric mountain waves from space. *Science*, **286**, 1534-1537.

Eckermann, S. D., and D. L. Wu, 2006: Imaging gravity waves in lower stratospheric AMSU-A radiances, Part 1: Simple forward model, *Atmos. Chem. Phys.*, **6**, 3325–3341.

Eckermann, S. D., D. L. Wu, J. D. Doyle, J. F. Burris, T. J. McGee, C. A. Hostetler, L. Coy, B. N. Lawrence, A. Stephens, J. P. McCormack, and T. F. Hogan, 2006: Imaging gravity waves in lower stratospheric AMSU-A radiances, Part 2: Validation case study. *Atmos. Chem. Phys.*, **6**, 3343-3362.

Eckermann, S. D., J. Ma, D. L. Wu, and D. Broutman, 2007: A three-dimensional mountain wave imaged in satellite radiance throughout the stratosphere: Evidence of the effects of directional wind shear. *Quart. J. Roy. Meteorol. Soc.*, **133**, 1959–1975.

Eckermann, S. D., L. Hoffmann, M. Höpfner, D. L. Wu and M. J. Alexander, 2009: Antarctic NAT PSC belt of June 2003: Observational validation of the mountain wave seeding hypothesis. *Geophys. Res. Lett.*, **36**, L02807, doi:10.1029/2008GL036629.

Ern, M., P. Preusse, and M. J. Alexander, 2004: Absolute values of gravity wave momentum flux derived from satellite data. *J. Geophys. Res.*, **109**, 10.1029/2004JD0004752.

Ern, M., P. Preusse, and C. D. Warner, 2006: Some experimental constraints for spectral parameters used in the Warner and McIntyre gravity wave parameterizations scheme. *Atmos. Chem. Phys.*, **6**, 4361-4381.

Ern M, Preusse P, Krebsbach M, Mlynczak MG, Russell III JM. 2008. Equatorial wave analysis from SABER and ECMWF temperatures. *Atmos. Chem. Phys.*, **8**, 845869.

Eskridge, R. E., O. A. Alduchov, I. V. Chernykh, Z. Panmao, A. C. Polansky and S. R. Doty, 1995: A comprehensive aerological reference data set (CARDS): Rough and systematic errors. *Bull. Am. Meteorol. Soc.*, **76**, 1759-1775.

Fetzer, E. J. and J. C. Gille, 1994: Gravity wave variance in LIMS temperatures. Part I: Variability and comparison with background winds. *J. Atmos. Sci.*, **51**, 2461-2483.

Fritts, D.C. and Alexander, M. J., 2003: Gravity wave dynamics and effects in the middle atmosphere. *Rev. Geophys.*, **41**, No.1, doi:10.1029/2001RG000106.

Fritts, D. C., S. L. Vadas, K. Wan, and J. A. Werne, 2006: Mean and variable forcing of the middle atmosphere by gravity waves. *J. Atmos. Solar-Terr. Phys.*, **68**, 247-265.

Fukao S., 2007: Recent Advances in Atmospheric Radar Study. *J. Meteor. Soc. Jpn.*, **85B**, 215-239.

Garcia, R. R. and B. A. Boville, 1994: "Downward Control" of the mean meridional circulation and temperature distribution of the polar winter stratosphere. *J. Atmos. Sci*, **51**, 2238-2245.

Garcia, R. R., D. R. Marsh, D. E. Kinnison, B. A. Boville, and F. Sassi, 2007: Simulation of secular trends in the middle atmosphere, 1950–2003. *J. Geophys. Res.*, **112**, 10.1029/2006JD007485.

Geller, M. A., and J. Gong, 2009: Gravity wave kinetic, potential, and vertical fluctuation energies as indicators of different frequency gravity waves. *J. Geophys. Res.*, (in press).

Giorgetta M. A., E. Manzini, and E. Roechner, 2002: Forcing of the quasi-biennial oscillation from a broad spectrum of atmospheric waves. *Geophysic. Res. Lett.* **29**, 1245, doi:10.1029/2002GL014756.

Giorgetta M. A., E. Manzini, E. Roechner, M. Esch, and L. Bengtsson, 2006: Climatology and forcing of the quasi-biennial oscillation in the MAECHAM5 model. *J. Climate*, **19**, 38823901.

Gong, J., M. A. Geller, and L. Wang, 2008: Source spectra information derived from U.S. high-resolution radiosonde data. *J. Geophys. Res.*, **113**, D10106, doi:10.1029/2007JD009252.

Grimsdell, A. W., M. J. Alexander, P. T. May, and L. Hoffmann, 2009: Model study of waves generated by convection with direct validation via satellite. *J. Atmos. Sci.*, (in press).

Hamilton, K. 1996: Comprehensive meteorological modelling of the middle atmosphere: A tutorial review, *J. Atmos. Terr. Phys.*, **35**,1591–1627.

Hamilton, K., and R. A. Vincent, 1995: High-resolution radiosonde data offer new prospects for research, *EOS*, **76**, 497.

Hamilton, K., R. J. Wilson, and R. Hemler, 1999: Atmosphere simulated with high vertical and horizontal resolution versions of a GCM:

Improvement in the cold pole bias and generation of a QBO-like oscillation in the tropics, *J. Atmos. Sci.* **56** 38293846.

Hamilton, K., R. J. Wilson, and R. Hemler, 2001: Spontaneous stratospheric QBO-like oscillations simulated by the GFDL SKYHI general circulation model, *J. Atmos. Sci.* **58**, 32713292.

Haynes, P., C. J. Marks, M. McIntyre, T. Shepherd, and K. Shine, 1991: On the "downward control" of extratropical diabatic circulations by eddy-induced mean zonal forces. *J. Atmos. Sci.*, **48**, 651–678.

Hecht, J. H., 2004: Instability layers and airglow imaging. *Rev. Geophys.*, **42**, RG1001, doi:10.1029/2003RG000131.

Hertzog, A. and Vial, F., 2001: A study of the dynamics of the equatorial lower stratosphere by use of ultra-long-duration balloons 2. Gravity waves. *J. Geophys. Res.*, **106**, 22,745-22,761.

Hertzog et al., 2007: Stratéole/Vorcore — Long-duration, superpressure balloons to study the Antarctic lower stratosphere during the 2005 winter. *J. Atmos. Ocean. Technol.*, **24**, 2048-2061.

Hertzog, A., G. Boccara, R. A. Vincent, F. Vial, and Ph. Cocquerez, 2008: Estimation of gravity-wave momentum fluxes and phase speeds from quasi-Lagrangian stratospheric balloon flights. 2: Results from the Vorcore campaign in Antarctica. *J. Atmos. Sci.*, **65**, 3056-3070.

Hines, C. O., 1997: Doppler-spread parameterization of gravity wave momentum deposition in the middle atmosphere, 2. Broad and quasi-monochromatic spectra, and implementation. *J. Atmos. Terr. Phys.*, **59**, 387400.

Hoffmann, L., and M. J. Alexander, 2009: Retrieval of stratospheric temperatures from Atmospheric Infrared Sounder radiance measurements for gravity wave studies. *J. Geophys. Res.*, **114**, D07105, doi:10.1029/2008JD011241.

Holton, J. R., 1982: The role of gravity wave induced drag and diffusion in the momentum budget of the mesosphere. *J. Atmos. Sci.*, **39**, 791–799.

Horinouchi T. and S. Yoden, 1998. Wave-mean flow interaction associated with a QBO-like oscillation simulated in a simplified GCM. *J. Atmos. Sci.*, **55**, 502526.

Jiang, J. H., S. D. Eckermann, D. L. Wu, and J. Ma, 2004. A search for mountain waves in MLS stratospheric limb radiances from the winter Northern Hemisphere: Data analysis and global mountain wave modeling. *J. Geophys. Res.*, **109**, D03107, doi:10.1029/2003JD003974.

Kawatani, Y., S. K. Dhaka, M. Takahashi, T. Tsuda, 2003: Large potential energy of gravity waves over a smooth surface with little convection. *Geophys. Res. Lett.*, **30**(8), 1438, doi:10.1029/2003GL016960.

Kawatani, Y., M. Takahashi, and T. Tokioka, 2004: Gravity waves around the subtropical jet of the southern winter in an atmospheric general circulation model. *Geophys. Res. Lett.*, **31**, L22109, doi:10.1029/2004GL020794.

Kawatani, Y., K. Tsuji, and M. Takahashi, 2005: Zonally non-uniform distribution of equatorial gravity waves in an atmospheric general circulation model. *Geophys. Res. Lett.*, **32**, L23815, doi:10.1029/2005GL024068.

Kawatani, Y., M. Takahashi, K. Sato, S. P. Alexander, and T. Tsuda, 2009: Global distribution of atmospheric waves in the equatorial upper troposphere and lower stratosphere: AGCM simulation of sources and propagation. *J. Geophys. Res.*, **114**, D01102. doi:10.1029/2008JD010374.

Kawatani, Y., K. Sato, T. J. Dunkerton, S. Watanabe, S. Miyahara, and M. Takahashi, 2010a: The roles of equatorial trapped waves and internal inertia-gravity waves in driving the quasi-biennial oscillation. Part I: zonal mean wave forcing. *J. Atmos. Sci.*, **67**, 963-980, DOI:10.1175/2009JAS3222.1.

Kawatani, Y., K. Sato, T. J. Dunkerton, S. Watanabe, S. Miyahara, and M. Takahashi, 2010b: The roles of equatorial trapped waves and internal inertia-gravity waves in driving the quasi-biennial oscillation. Part II: Three-dimensional distribution of wave forcing. *J. Atmos. Sci.*, **67**, 981-997, DOI:10.1175/2009JAS3223.1.

Koshyk, J. N., B. A. Boville, K. Hamilton, E. Manzini, and K. Shibata, 1999: Kinetic energy spectrum of horizontal motions in middle-atmosphere models. *J. Geophys. Res.*, **104** 27177-27190.

Lawrence, B. A., 1997: Some aspects of the sensitivity of stratospheric climate simulations to model lid height. *J. Geophys. Res.*, **102**, 2380523811.

Li, F., J. Austin, and J. Wilson, 2008: The strength of the Brewer-Dobson circulation in a changing climate: coupled chemistry-climate model simulations. *J. Clim.*, **21**, 40–57.

Limpasuvan, V., D. L. Wu, M. Joan Alexander, M. Xue, M. Hu, S. Pawson, and J. R. Perkins, 2007: Stratospheric gravity wave simulation over Greenland during 24 January 2005. *J. Geophys. Res.*, **112**, D10115, doi:10.1029/2006JD007823.

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

Manzini, E., B. Steil, C. Brühl, M. A. Giorgetta, and K. Krüger, 2003: A new interactive chemistry-climate model: 2. Sensitivity of the middle atmosphere to ozone depletion and increase in greenhouse gases and implications for recent stratospheric cooling. *J. Geophys. Res.*, **108**, 10.1029/2002JD002977.

Massman, W. J., 1981: An investigation of gravity waves on a global scale using TWERLE data, *J. Geophys. Res.*, **86**, 4072-4082.

Marks, C. J. 1989: Some features of the climatology of the middle atmosphere revealed by Nimbus 5 and 6. *J. Atmos. Sci.*, **46**, 24852508.

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

McLandress, C., 1998: On the importance of gravity waves in the middle atmosphere and their parameterization in general circulation models. *J. Atmos. Sol.–Terr. Phys.*, **60**, 1357–1383.

McLandress, C., and T. G. Shepherd, 2009: Simulated anthropogenic changes in the Brewer-Dobson circulation, including its extension to high latitudes. *J. Clim.*, **22**, 1516–1540.

Mende, S. B., G. R. Swenson, S. P. Geller, and K. A. Spear, 1994: Topside observation of gravity waves. *Geophys. Res. Lett.*, **21**, 2283–2286.

Miyahara, S., 2006: A three-dimensional wave activity flux applicable to inertio-gravity waves. *SOLA*, **2**, 108111.

Miyahara, S., Y. Hayashi, and J. D. Mahlman, 1986: Interactions between gravity waves and planetary scale flow simulated by the GFDL SKYHI general circulation model. *J. Atmos. Sci.*, **43**, 18441861.

Nastrom, G. D., 1980: The response of superpressure balloons to gravity waves. *J. Appl. Meteor.*, **19**, 1013-1019.

Nezlin, Y., Y. J. Rochon, S. Polavarapu, 2009: Impact of tropospheric and stratospheric data assimilation on mesosphere prediction. *Tellus*, **61A**, 154-159.

O'Sullivan, D., and T. J. Dunkerton, 1995: Generation of inertia-gravity waves in a simulated life cycle of baroclinic instability. *J. Atmos. Sci.*, **52**, 3695-3716.

Palmer, T., G. Shutts, R. Swinbank, 1986: Alleviation of systematic westerly bias in general circulation and numerical weather prediction models through an orographic gravity wave drag parameterization. *Quart. J. Roy. Met. Soc.*, **112**, 2056–2066.

Parkinson, C. L., 2003: Aqua: An Earth-observing satellite mission to examine water and other climate variables. *IEEE Trans. Geosci. Remote Sensing*, **41**, 173–183.

Picard, R. H., R. R. O'Neil, H. A. Gardiner, J. Gibson, J. R. Winick, W. O. Gallery, A. T. Stair Jr., P. P. Wintersteiner, E. R. Hegblom, and E. Richards, 1998: Remote sensing of discrete stratospheric gravity-wave structure at 4.3-μm from the MSX satellite, *Geophys. Res. Lett.*, **25**, 2809–2812.

Plougonven, R. and C. Snyder, 2007: InertiaGravity Waves Spontaneously Generated by Jets and Fronts. Part I: Different Baroclinic Life Cycles. *J. Atmos. Sci.*, **64**, 2502-2520.

Preusse P., A. Dörnbrack, S. D. Eckermann, M. Riese, B. Schaeler, J. T. Bacmeister, D. Broutman, and K. U. Grossmann. 2002. Space-based measurements of stratospheric mountain waves by CRISTA, 1. Sensitivity, analysis method, and a case study. *J. Geophys. Res.*, **107**(D23), 8178, doi:10.1029/2001JD000699.

Preusse, P., S. D. Eckermann, and M. Ern, 2008: Transparency of the atmosphere to short horizontal wavelength gravity waves. *J. Geophys. Res.*, **113**, D24104, doi:10.1029/2007JD009682.

Pulido, M. and J. Thuburn, 2005: Gravity wave drag estimation from global analyses using variational data assimilation principles. I: Theory and implementation. *Q. J. Roy. Meteorol. Soc.*, **131**, 1821-1840.

Pulido M. and J. Thuburn, 2006: Gravity wave drag estimation from global analyses using variational data assimilation principles. II: A case study. *Q. J. Roy. Meteorol. Soc.* , **132**, 1527-1543. DOI: 10.1256/qj.05.43

Pulido M. and J. Thuburn, 2008: The seasonal cycle of gravity wave drag in the middle atmosphere. *J. Climate*, **21**, 4664-4679.

Ren, S., S. Polavarapu and T. G. Shepherd, 2008: Vertical propagation of information in a middle atmosphere data assimilation system. *Geophys. Res. Let.*, **35**, L06804, doi:10.1029/2007GL032699.

Richter, J. H., F. Sassi, R. R. Garcia, K. Matthes, and C. A. Fischer, 2008: Dynamics of the middle atmosphere as simulated by the Whole Atmosphere Community Climate Model, version 3 (WACCM3). *J. Geophys. Res.*, **113**, 10.1029/2007JD009269.

Richter, J. H., F. Sassi, R. R. Garcia, 2010: Toward a physically based gravity wave source parameterization in a general circulation model. *J. Atmos. Sci.*, **67**, 136-156.

Riggin, D M., D. C. Fritts, C. D. Fawcett, E. Kudeki, and M. H. Hitchman, 1997:Radar observations of gravity waves over Jicamarca, Peru during the CADRE campaign. *J. Geophys. Res.*, **102**, 26,263-26,281.

Ryu, J.-H., S. Lee, and S.-W. Son, 2008: Vertically propagating Kelvin waves and tropical tropopause variability. *J. Atmos. Sci.*, **65**, 1817-1837.

Salby, M. L., 1982: Sampling theory for asynoptic satellite observations. Part II: Fast Fourier synoptic mapping. *J. Atmos. Sci.*, **39**, 2601-2614.

Sato, K. 1993: Small-scale wind disturbances observed by the MU radar during the passage of typhoon Kelly. *J. Atmos. Sci.*, **50**, 518-537.

Sato, K. 1994: A statistical study of the structure, saturation and sources of inertio-gravity waves in the lower stratosphere observed with the MU radar. *J. Atmos. Terr. Phys.*, **56**, 755-774.

Sato, K. 2000: Sources of gravity waves in the polar middle atmosphere. *Adv. Polar Upper Atmos. Res.*, **14**, 233240.

Sato, K., T. Kumakura, and M. Takahashi, 1999: Gravity waves appearing in a high-resolution GCM simulation. *J. Atmos. Sci.*, **56**, 1005-1018.

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

Sato, K., S. Watanabe, Y. Kawatani, Y. Tomikawa, K. Miyazaki, and M. Takahashi, 2009a: On the origins of gravity waves in the mesosphere. *Geophys. Res. Lett.*, **36**, L19801, doi:10.1029/2009GL039908.

Sato, K., S. Tateno, S. Watanabe, and Y. Kawatani, 2009b: Gravity

waves in the middle to high latitudes of the Southern Hemisphere using a high-resolution GCM. *J. Atmos. Sci.*, (accepted).

Scinocca, J. F., N. A. McFarlane, M. Lazare, J. Li, and D. Plummer, 2008: Technical Note: The CCCma third generation AGCM and its extension into the middle atmosphere. *Atmos. Chem. Phys.*, **8**, 7055–7074.

Scinocca, J. F., and N. A. McFarlane, 2000: The parameterization of drag induced by stratified flow over anistropic orography. *Quart J. Roy. Met. Soc.*, **126**, 2353–2393.

Shaw, T. A., and T. G. Shepherd, 2007: Angular momentum conservation and gravity wave drag parameterization: Implications for climate models. *J. Atmos. Sci.*, **64**, 190203.

Shaw, T. A., M. Sigmond, T. G. Shepherd, and J. F. Scinocca, 2009: Sensitivity of simulated climate to conservation of momentum in gravity wave drag parameterization. *J. Clim.*, (in press).

Shepherd, T. G., and T. A. Shaw, 2004: The angular momentum constraint on climate sensitivity and downward influence in the middle atmosphere. *J. Atmos. Sci.*, **61**, 28992908.

Shibata, K. and M. Deushi, 2005: Partitioning between resolved wave forcing and unresolved gravity wave forcing to the quasi-biennial oscillation as revealed with a coupled chemistry-climate model. *Geophys. Res. Lett.*, **32**, L12820, doi:10.1029/2005GL022885.

Shine, K. P., 1989: Sources and sinks of zonal momentum in the middle atmosphere diagnosed using the diabatic circulation. *Q. J. R. Meteorol. Soc.*, **115**, 265292.

Shutts, G., 1995: Gravity-wave drag prametrization over complex terrain: The effect of critical-level absorption in directional wind-shear. *Q. J. Roy. Meteorol. Soc.*, **121**, 1005-1021.

Sigmond, M., P. J. Kushner, and J. F. Scinocca, 2007: Discriminating robust and non-robust atmospheric circulation responses to global warming. *J. Geophys. Res.*, **112**, doi:10.1029/2006JD008270.

Sigmond, M. and J. F. Scinocca, 2010: The influence of the basic state on the Northern Hemisphere circulation response to climate change, *J.Clim.*, **23**, 1434-1446.

Sigmond, M., J. F. Scinocca, and P. J. Kushner, 2008: The impact of the stratosphere on tropospheric climate change. *Geophys. Res. Lett.*, **35**, L12, 706.

Siskind, D. E., S. D. Eckermann, L. Coy, J. P. McCormack, and C. E. Randall, 2007: On recent interannual variability of the Arctic winter mesosphere: Implications for tracer descent. *Geophys. Res. Lett.*, **34**, L09806, doi:10.1029/2007GL029293.

Solomon, S. 1988: The mystery of the Antarctic ozone "hole". *Rev. Geophys.*, **26**, 131-148.

Song I.-S, and H.-Y. Chun. 2008. A Lagrangian spectral parameterization of gravity wave drag induced by cumulus convection. *J. Atmos. Sci.*, **65**, 1204-1224.

Song, Y. and W. A. Robinson, 2004: Dynamical mechanisms for stratospheric influences on the troposphere. *J. Atmos. Sci.*, **61**, 1711–1725.

Sugimoto, N., K. Ishioka, and K. Ishii, 2008: Parameter sweep experiments on spontaneous gravity wave radiation from unsteady rotational flow in an f-plane shallow water system. *J. Atmos. Sci.*, **65**, 235-249. doi: 10.1175/2007JAS2404.1.

Susskind, J., C. Barnet, J. Blaisdell, L. Iredell, F. Keita, L. Kouvaris, G. Molnar, and M. Chahine, 2006: Accuracy of geophysical parameters derived from Atmospheric Infrared Sounder/Advanced Microwave Sounding Unit as a function of fractional cloud cover. *J. Geophys. Res.*, **111**, D09S17, doi:10.1029/2005JD006272.

Takahashi, M., 1996: Simulation of the stratospheric quasi-biennial oscillation using a general circulation model. *Geophys. Res. Lett.*, **23**, 661664.

Takahashi, M., 1999: Simulation of the quasi-biennial oscillation in a general circulation model. *Geophys. Res. Lett.*, **26**, 13071310.

Tateno, S. and K. Sato, 2008: A study of inertia-gravity waves in the middle stratosphere based on intensive radiosonde observations. *J. Meteor. Soc. Japan.*, **85** 719-732.

Tsuda, T., Y. Murayama, M. Yamamoto, S. Kato, S. Fukao, 1990: Seasonal variation of momentum fluxes in the mesosphere observed with the MU radar. *Geophys. Res. Lett.*, **17**, 725-728.

The TWERLE team, 1977: The TWERL Experiment. *Bull. Am. Meteorol. Soc.*, **58**, 936-948.

Vincent, R. A., and I. M. Reid, 1983: HF Doppler measurements of mesospheric gravity wave momentum fluxes. *J. Atmos. Sci.*, **40**, 13211333.

Vincent, R. A., S. J. Allen, and S. D. Eckermann, 1997: Gravity-wave parameters in the lower atmosphere. in *Gravity Wave Processes: Their Parameterization in Global Climate Models*, K. Hamilton, Ed., Springer-Verlag, 7-25.

Vincent, R. A., A. Hertzog, G. Boccara, and F. Vial, 2007: Quasi-Lagrangian superpressure balloon measurements of gravity-wave momentum fluxes in the polar stratosphere of both hemispheres, *Geophys. Res. Lett.*, **34**, L19804, doi:10.1029/2007GL031072.

Wang, L., 2003: Gravity wave analysis of four years of high vertical resolution U. S. radiosonde data. PhD dissertation, Marine and Atmospheric Sciences, Stony Brook University, 130 pp.

Wang, L., Geller, M. A., and M. J. Alexander, 2005: Spatial and temporal variations of gravity wave parameters. Part I: Intrinsic frequency, wavelength, and vertical propagation direction. *J. Atmos. Sci.*, **62**, 125-142.

Watanabe, S., 2008: Constraints on a non-orographic gravity wave drag parameterization using a gravity wave resolving general circulation

model. *Sci. Online Lett. Atmos.*, **4**, 061-064, doi:10.2151/sola.2008-016.

Watanabe, S. and S. Miyahara, 2009: Quantification of the gravity wave forcing of the migrating diurnal tide in a gravity wave-resolving general circulation model. *J. Geophys. Res.*, **114**, D07110, doi:10.1029/2008JD011218.

Watanabe, S., K. Sato, and M. Takahashi, 2006: A general circulation model study of the orographic gravity waves over Antarctica excited by katabatic winds. *J. Geophys. Res.*, **111**, D18104, doi:10.1029/2005JD006851.

Watanabe, S., Y. Kawatani, Y. Tomikawa, K. Miyazaki, M. Takahashi, and K. Sato, 2008: General aspects of a T213L256 middle atmosphere general circulation model. *J. Geophys. Res.*, **113**, D12110, doi:10.1029/2008JD010026.

Watanabe, S., Y. Tomikawa, K. Sato, Y. Kawatani, K. Miyazaki, and M. Takahashi, 2009: Simulation of the eastward 4-day wave in the Antarctic winter mesosphere using a gravity wave resolving general circulation model. *J. Geophys. Res.*, doi:10.1029/2008JD011636, (in press).

Wu, D. L., 2004: Mesoscale gravity wave variances from AMSU-A radiances, *Geophys. Res. Lett.*, **31**, L12114, doi:10.1029/2004GL019562.

Wu, D. L., and J. W. Waters, 1996: Satellite observations of atmospheric variances: A possible indication of gravity waves. *Geophys. Res. Lett.*, **23**, 631-3634.

Wu, D. L., and F. Zhang, 2004: A study of mesoscale gravity waves over the North Atlantic with satellite observations and a mesoscale model. *J. Geophys. Res.*, **109**, D22104, doi:10.1029/2004JD005090.

Wu, D. L., P. Preusse, S. D. Eckermann, J. H. Jiang, M. de la Torre Juarez, L. Coy, and D. Y. Wang, 2006: Remote sounding of atmospheric gravity waves with satellite limb and nadir techniques. *Adv. Space Res.*, **37**, 2269–2277.

Zhang, F., 2004: Generation of mesoscale gravity waves in upper-tropospheric jetfront systems. *J. Atmos. Sci.*, **61**, 440-457.

