Driving of the SAO by gravity waves as observed from satellite

M. Ern, P. Preusse, and M. Riese
Institut für Energie- und Klimaforschung – Stratosphäre (IEK–7), Forschungszentrum Jülich GmbH, 52425 Jülich, Germany

Correspondence to: M. Ern (m.ern@fz-juelich.de)

Received: 10 October 2014 – Revised: 2 April 2015 – Accepted: 7 April 2015 – Published: 29 April 2015

Abstract. It is known that atmospheric dynamics in the tropical stratosphere have an influence on higher altitudes and latitudes as well as on surface weather and climate. In the tropics, the dynamics are governed by an interplay of the quasi-biennial oscillation (QBO) and semiannual oscillation (SAO) of the zonal wind. The QBO is dominant in the lower and middle stratosphere, and the SAO in the upper stratosphere/lower mesosphere. For both QBO and SAO the driving by atmospheric waves plays an important role. In particular, the role of gravity waves is still not well understood.

In our study we use observations of the High Resolution Dynamics Limb Sounder (HIRDLS) satellite instrument to derive gravity wave momentum fluxes and gravity wave drag in order to investigate the interaction of gravity waves with the SAO. These observations are compared with the ERA-Interim reanalysis. Usually, QBO westward winds are much stronger than QBO eastward winds. Therefore, mainly gravity waves with westward-directed phase speeds are filtered out through critical-level filtering already below the stratopause region. Accordingly, HIRDLS observations show that gravity waves contribute to the SAO momentum budget mainly during eastward wind shear, and not much during westward wind shear. These findings confirm theoretical expectations and are qualitatively in good agreement with ERA-Interim and other modeling studies. In ERA-Interim most of the westward SAO driving is due to planetary waves, likely of extratropical origin. Still, we find in both observations and ERA-Interim that sometimes westward-propagating gravity waves may contribute to the westward driving of the SAO. Four characteristic cases of atmospheric background conditions are identified. The forcings of the SAO in these cases are discussed in detail, supported by gravity wave spectra observed by HIRDLS. In particular, we find that the gravity wave forcing of the SAO cannot be explained by critical-level filtering alone; gravity wave saturation without critical levels being reached is also important.

Keywords. Meteorology and atmospheric dynamics (general circulation; middle atmosphere dynamics; waves and tides)

1 Introduction

In the tropical stratosphere and lower mesosphere, the zonal wind is dominated by an interplay of the quasi-biennial oscillation (QBO) in the lower and middle stratosphere and the semiannual oscillation (SAO) in the upper stratosphere/lower mesosphere. The QBO has an average period of 28 months. Usually the QBO winds are asymmetric with a strong westward wind phase (as strong as about $-40\text{ m s}^{-1}$) and much weaker eastward winds (only about $20\text{ m s}^{-1}$ at maximum).

The SAO has a period of 6 months, and both eastward and westward wind can be quite strong: about $-60$ to $-20\text{ m s}^{-1}$ for westward wind, and about $20$ to $40\text{ m s}^{-1}$ for eastward wind (e.g., Hirota, 1980; Delisi and Dunkerton, 1988; Dee et al., 2011, and references therein). More details about QBO and SAO can be found in Baldwin et al. (2001) and references therein.

The QBO and the SAO are important processes in atmospheric dynamics. Both QBO and SAO have an effect on the tracer transport in the stratosphere (e.g., Huang et al., 2008; Punge et al., 2009; Khosrawi et al., 2013; Shu et al., 2013). Further, the QBO has an effect on the stability of the polar vortex (e.g., Holton and Tan, 1980), and there are indications that both QBO and SAO have an influence on the timing of sudden stratospheric warmings (e.g., Pascoe et al., 2006). It has been found that the QBO has an effect on the weather and climate in the lower atmosphere and even at the surface (e.g., Ebden, 1975; Boer and Hamilton, 2008; Marshall and Scaife, 2009).

Because of their importance, the tropics have been the focus of previous measurement campaigns and will be the

Published by Copernicus Publications on behalf of the European Geosciences Union.
topic of future ones (e.g., Deepa et al., 2006; Cairo et al., 2010; Jewtoukoff et al., 2013; von Hobe et al., 2013). In addition, modeling efforts are currently underway to improve the representation of the tropics and, in particular, the QBO in weather and climate models (e.g., Yang et al., 2009; Kim et al., 2013; Kim and Chun, 2015; Lott et al., 2012, 2014; Krismer and Giorgetta, 2014; Maury and Lott, 2014; Richter et al., 2014). With a more realistic model representation of the QBO, potentially the coupling toward higher latitudes and even seasonal weather prediction might be improved (e.g., Gerber et al., 2012; Scaife et al., 2014).

Both QBO and SAO filter the spectrum of waves that propagate upward. This filtering of waves is relevant for the formation of circulation patterns at higher altitudes. For example, the pre-filtered wave spectrum is likely responsible for the formation of a QBO and an SAO in the tropical mesopause region (see also Dunkerton, 1982; Burragate et al., 1996; Baldwin et al., 2001; Richter and Garcia, 2006; Peña-Ortiz et al., 2010; de Wit et al., 2013; Kishore Kumar et al., 2014, and references therein). It has also been found that the QBO and the SAO interact with each other. For example, the QBO and SAO periods are often synchronized (e.g., Dunkerton and Delisi, 1997; Krismer et al., 2013), and it has been suggested that the eastward phase of the SAO can initiate an eastward phase of the QBO (e.g., Kuai et al., 2009).

It was proposed by Lindzen and Holton (1968) and Holton and Lindzen (1972) that the QBO is a wave-driven circulation. The wave driving by planetary waves alone is, however, not sufficient, and it has been concluded that most of the wave driving is contributed by mesoscale gravity waves (e.g., Dunkerton, 1997; Ern and Preusse, 2009a, b). This is also in agreement with direct observations of the QBO driving by gravity waves (Ern et al., 2014).

Similarly, the wind reversal of the SAO from westward to eastward winds is likely driven by gravity waves and (to a minor extent) by planetary waves, for example equatorially trapped Kelvin waves. Different from this, the wind reversal from SAO eastward to westward wind is assumed to be mainly driven by horizontal advection and meridional momentum transport of extratropical planetary waves (e.g., Reed, 1966; Hirota, 1980; Delisi and Dunkerton, 1988; Hamilton and Mahlmann, 1988). Therefore the descent of the SAO westward wind phase with time is usually much steeper than the descent of the SAO eastward wind phase, which is mainly driven by vertically propagating waves. In particular, several studies suggest that the filtering of the spectrum of upward-propagating waves by the QBO has a strong influence on the SAO winds (e.g., Hirota, 1980; Garcia et al., 1997; Peña-Ortiz et al., 2010). This wave-filtering effect of the QBO (mainly critical-level filtering) is also seen in observed gravity wave momentum flux spectra (Ern et al., 2014).

A number of general circulation models (GCMs) and chemistry–climate models (CCMs) are able to simulate an SAO. In most simulations, the SAO is driven by a combination of resolved waves and parameterization of subgrid-scale gravity waves (e.g., Orr et al., 2010; Peña-Ortiz et al., 2010; Osprey et al., 2010). In some simulations, the SAO is driven alone by gravity waves and planetary waves explicitly resolved by the model (e.g., Kawatani et al., 2010).

The role of the different terms in the tropical momentum balance, in particular the role of gravity wave drag, is, however, strongly dependent on the model setup and model resolution (e.g., Osprey et al., 2010). This shows that there is still large uncertainty about details of the forcing of the SAO.

To obtain a more realistic QBO and SAO in GCMs/CCMs, an improvement in the parameterized gravity wave drive is required. Up to now most observational estimates of the gravity wave contribution to the SAO momentum budget have been from ground-based stations (e.g., Deepa et al., 2006; Antonita et al., 2007). In order to constrain gravity wave parameterizations, however, global observations (from satellite) are needed (e.g., Ern et al., 2006; Alexander et al., 2010; Geller et al., 2013). Several previous studies based on global satellite observations indicate that the gravity wave distribution in the tropics interacts with the QBO and SAO winds (e.g., de la Torre et al., 2006; Krebsbach and Preusse, 2007; Wu and Eckermann, 2008). These studies were, however, limited to gravity wave variances or squared amplitudes. It was only recently that information about gravity wave momentum fluxes, gravity wave drag, and detailed spectral information were obtained for the QBO by Ern et al. (2014). For the SAO, the direct estimation of gravity wave drag from global observations is still an open issue. Further, spectral information about the gravity waves that contribute to the SAO can help to improve our physical understanding of the wave dynamics in the tropics.

In our study we use satellite observations of gravity waves by the High Resolution Dynamics Limb Sounder (HIRDLS) instrument to investigate how gravity waves contribute to the driving of the SAO. In Sect. 2.1 some information about the HIRDLS instrument is given, as well as descriptions of how gravity waves are extracted from the temperature observations and how gravity wave momentum fluxes and drag are derived. In Sect. 3 we determine the SAO momentum budget from the European Centre for Medium-Range Weather Forecasts (ECMWF) ERA-Interim reanalysis. In Sect. 4 it is shown how HIRDLS gravity wave variances and momentum fluxes are modulated by the SAO. Further, gravity wave drag is calculated from the momentum fluxes and compared to the SAO momentum budget in ERA-Interim. The driving of the SAO by gravity waves is discussed in detail for four characteristic cases in Sect. 5. In Sect. 6 this discussion is supported by presenting gravity wave spectra that are determined from the observations. Finally, in Sect. 7 the results are summarized and discussed.
2 Data and methods

2.1 Satellite data and related gravity wave diagnostics

Our study is based on temperature observations of the HIRDLS satellite instrument and the gravity wave momentum fluxes that are estimated from these observations. In the following, some information about the HIRDLS instrument is given, and the procedure to derive gravity wave momentum fluxes and drag is introduced.

2.1.1 The HIRDLS instrument

The HIRDLS instrument observes atmospheric limb emissions of CO$_2$ at 15 µm in the infrared. From these observations, altitude profiles of atmospheric temperature are derived, as well as several trace species. HIRDLS is onboard the EOS-Aura satellite, and temperature observations are available from January 2005 until March 2008. The altitude range covered is from the tropopause region to the upper mesosphere. The vertical resolution of the observed temperature altitude profiles is close to the vertical field of view of the instrument (about 1 km). The HIRDLS horizontal sampling distance between consecutive altitude profiles is about 90 km.

More information about the HIRDLS instrument and temperature retrieval is given, for example, in Gille et al. (2003, 2008) and Barnett et al. (2008). In our study we use HIRDLS V006 temperatures (see also Gille et al., 2011).

2.1.2 Extraction of gravity waves, and estimation of gravity wave momentum flux and drag

To investigate the role of gravity waves in the forcing of the SAO we derive gravity wave variances, momentum fluxes and drag from HIRDLS temperature observations. All time series presented later in our work are averages over the latitude band 10° S–10° N and over 7 days with a time step of 3 days. This provides both good statistics and a time resolution that is sufficient to sample the rapid circulation changes that are associated with the SAO.

Extraction of gravity waves

In order to extract gravity waves from satellite temperature observations we follow the procedure described in Ern et al. (2011).

In the first step, from observed temperature altitude profiles the zonal-mean background temperature is subtracted, as well as stationary and traveling global-scale waves of zonal wave numbers 1–6. In particular, Kelvin waves in the tropics, which can have very short periods of only a few days (e.g., Smith et al., 2002; Forbes et al., 2009; Ern et al., 2009a), are also removed. The strongest tidal modes are removed by subtracting quasi-stationary zonal wave numbers 0–4 separately for ascending and descending parts of the satellite orbits (see also Ern et al., 2013).

The result after this first step are altitude profiles of residual temperatures that can be attributed to mesoscale gravity waves. The strongest vertical wave structures in these altitude profiles are determined by a two-stage method called MEM/HA, which is described in detail in Preusse et al. (2002). The result is vertical profiles of wave amplitudes, vertical wavelengths and vertical phases of the strongest gravity waves for each altitude profile of residual temperatures. In our study, these gravity wave parameters are determined in windows of 10 km vertical extent (see also Ern et al., 2004, 2011). In this way a large vertical wavelength range of 2–25 km is covered by the analysis.

Estimation of absolute momentum fluxes

The absolute momentum flux $F_{ph}$ carried by an observed gravity wave is calculated using the following equation (Ern et al., 2004):

$$
F_{ph} = \frac{1}{2} \varrho_0 \frac{\lambda_h}{\lambda_z} \left( \frac{g}{N} \right)^2 \left( \frac{\hat{T}}{T} \right)^2.
$$

(1)

In this equation $\lambda_h$ and $\lambda_z$ are the horizontal and vertical wavelength of the gravity wave, $\varrho_0$ is the atmospheric density, $g$ the gravity acceleration, $N$ the buoyancy frequency, $\hat{T}$ the temperature amplitude of the wave, and $T$ the atmospheric background temperature.

The horizontal wavelength of a gravity wave is determined from pairs of altitude profiles. Like in Ern et al. (2011), we assume that the same wave is observed in two consecutive altitude profiles of a given satellite measurement track (profile pairs) if the vertical wavelengths in these altitude profiles differ by no more than 40%. In the tropics this is the case for about 60–70% of all profile pairs. The HIRDLS along-track sampling time step is about 10–15 s on average. Therefore, it can be assumed that a gravity wave is observed quasi-instantaneously by two consecutive altitude profiles (see also Ern et al., 2011). The horizontal wavelength of the observed gravity wave is estimated from the shift of the vertical phase of the wave between the two altitude profiles (see also Ern et al., 2004, 2011). This horizontal wavelength is, however, only the projection of the true horizontal wavelength of the gravity wave on the satellite measurement track, and is therefore always an overestimation. See also the discussion in Preusse et al. (2009).

Because the spatial orientation of the observed gravity wave cannot be determined from a single satellite measurement track, no directional information is available, and the momentum fluxes are only absolute (total) momentum fluxes. See also the detailed discussion in Ern et al. (2014).

The uncertainty of these total momentum fluxes is large, at least a factor of 2. Two main error sources are uncertainties in the horizontal wavelength and the sensitivity of the
instrument for the detection of gravity waves. This sensitivity decreases close to the detection limits at short horizontal and short vertical wavelengths. Satellite instruments observing in limb-viewing geometry can only detect gravity waves with horizontal wavelengths \(> 100–200\) km (see also Preusse et al., 2002, 2009; Ungermann et al., 2010), and therefore observe only part of the momentum flux of the whole spectrum of gravity waves (e.g., Ern et al., 2005; Choi et al., 2012). Further, the vertical resolution of an instrument limits the range of vertical wavelengths that can be detected. Therefore HIRDLS is sensitive only to gravity waves with vertical wavelengths \(> 2\) km. For a more detailed error discussion see also Ern et al. (2004) and Trinh et al. (2015).

Estimation of total (absolute) gravity wave drag

The total (absolute) drag \(XY\) by gravity waves on the background flow can be calculated from vertical gradients of total (absolute) momentum flux:

\[
XY = -\frac{1}{\rho_0} \frac{\partial F_{ph}}{\partial z},
\]

with \(z\) the vertical coordinate. Because the total momentum flux \(F_{ph}\) is only an absolute value, the total drag \(XY\) calculated from its vertical gradient also contains no directional information. Like total momentum fluxes, total gravity wave drag has large uncertainties, at least a factor of 2. The net gravity wave drag that is exerted on the background flow due to the single gravity waves in a certain region cancels out (see also Geller et al., 2013; Ern et al., 2011; Ern et al., 2014).

This lack of directional information can, however, often be overcome. For example, for the case of a wind reversal at the top of a strong wind jet, it can be assumed that the momentum flux distribution below the wind reversal is dominated by gravity waves propagating opposite to the wind direction in the jet (e.g., Warner et al., 2005; Gong et al., 2008). If these waves, while propagating upward, encounter the wind reversal at the top of the jet, they will dissipate more strongly because they either encounter critical wind levels or their propagation conditions become less favorable (intrinsic frequencies and thus critical amplitudes for the onset of wave breaking are reduced). For such cases the resulting net drag will then be close to equal to the total drag and will be opposite to the prevailing wind direction in the jet (see also Warner et al., 2005). Cases where the assumption of a prevalent direction of the total gravity wave drag observed from satellite led to meaningful results are, for example, the reversal of the summertime mesospheric jet (Ern et al., 2013) or the gravity wave driving of the QBO (Ern et al., 2014).

In our current study the situation is sometimes more complicated because the spectrum of gravity waves that reaches the stratopause region and contributes to the driving of the SAO is pre-filtered by the QBO in the lower and middle stratosphere. This is discussed in detail in Sects. 4–6.

2.2 ERA-Interim and the TEM zonal momentum budget

As mentioned in the Introduction, the contribution of gravity waves in the tropical momentum budget of the SAO is not well known and differs between different simulations. Therefore, one of the main goals of our study is to provide some guidance for global models regarding the role of gravity waves in the SAO momentum budget. For this purpose, realistic background winds are required for the period and altitude range considered. Previous studies have shown that zonal winds in the tropics provided by ECMWF are in good agreement with observations (e.g., Baldwin and Gray, 2005; Dee et al., 2011; Baron et al., 2013). In particular, there is qualitatively good agreement between the QBO-related gravity wave drag variations derived from the ECMWF ERA-Interim reanalysis and those derived from satellite observations (Ern et al., 2014). Therefore we use ERA-Interim also for studying the role of gravity waves in the driving of the SAO. More information about ERA-Interim can be found in Dee et al. (2011). For our study, ERA-Interim data are interpolated on a horizontal longitude/latitude grid of \(1^\circ \times 1^\circ\) resolution. The vertical resolution is about \(1.4\) km. Further, we use all available time steps (00:00, 06:00, 12:00 and 18:00 GMT) to avoid biases by diurnal cycle effects (see also Seviour et al., 2012).

The different terms of the ERA-Interim momentum budget are calculated for each of the 6 h time steps. In order to match the temporal resolution of our HIRDLS gravity wave data, 7-day averages are calculated from these single estimates every 3 days (see also Sect. 2.1.2). Finally, latitudinal averages are calculated over the tropical latitude band \(10^\circ \text{S}–10^\circ \text{N}\).

As detailed in Andrews et al. (1987), the transformed Eulerian mean (TEM) momentum budget of the zonal-mean zonal wind can be written as follows:

\[
\frac{\partial \overline{u}}{\partial t} + \overline{u} \cdot \nabla \phi + \frac{\overline{u}}{a \cos \phi} - f = \overline{w^* z} = \overline{F}_{PW} + \overline{F}_{GW},
\]

with \(\overline{u}\) the zonal-mean zonal wind, \(\partial \overline{u}/\partial t\) its tendency, and \(\nabla \phi\) and \(\overline{w^* z}\) the TEM meridional and vertical wind, respectively. Further, \(f\) is the Coriolis frequency, \(a\) the Earth’s radius, and \(\phi\) the geographic latitude. \(\overline{F}_{PW}\) and \(\overline{F}_{GW}\) are the zonal-mean wave drag due to planetary waves and gravity waves, respectively. Subscripts \(\phi\) and \(z\) stand for differentiation in meridional and vertical direction, respectively, and overbars indicate zonal averages. In the following, the momentum terms involving \(\nabla \phi\) and \(\overline{w^* z}\) will be called “meridional advection term” and “vertical advection term”, respectively.

Generally, the drag of resolved waves \(\overline{F}_{\phi z}\) can be calculated from the divergence of the Eliassen–Palm flux (EP flux). The meridional \((F_{\phi})\) and vertical component \((F^{(z)})\)
usually, in models the contribution of gravity waves \( \overline{X}_{GW} \) comprises the contribution of all resolved waves with higher zonal wave numbers \( (k > 20) \), gravity wave drag \( \overline{X}_{\text{param}} \) that is simulated by dedicated parameterizations (in the case of ERA-Interim just by Rayleigh friction), and the remaining imbalance \( \overline{X}_{\text{imbalance}} \) in the momentum budget that is introduced, for example, by data assimilation. In our case, the overall contribution of gravity waves can be written as

\[
\overline{X}_{GW} = \overline{X}_{\text{res}}(k > 20) + \overline{X}_{\text{param}} + \overline{X}_{\text{imbalance}}.
\]

In the ECMWF model, the contribution \( \overline{X}_{\text{res}}(k > 20) \), attributed to gravity waves resolved by the model, severely underestimates the contribution of gravity waves in the real atmosphere, and the distribution of resolved waves of high zonal wave number in the tropics is not very realistic (e.g., Schroeder et al., 2009; Preusse et al., 2014). Further, \( \overline{X}_{\text{param}} \) and \( \overline{X}_{\text{imbalance}} \) are not standard model output and are therefore not known. For this reason, we follow an approach similar to the one presented in Alexander and Rosenlof (1996) and estimate the contribution \( \overline{X}_{GW} \) in Eq. (3) indirectly. This is done by calculating \( \overline{X}_{GW} \) in Eq. (3) as “missing drag” from all the other terms in the ERA-Interim momentum budget, thereby assuming that all other contributions in Eq. (3) are known and realistic (see also Ern et al., 2014). Given a good underlying model and by assimilating a considerable amount of data, this missing drag can be assumed to be the contribution of gravity wave drag in the zonal momentum budget (e.g., Alexander and Rosenlof, 1996; Ern et al., 2014). Even if the missing drag should no longer be fully reliable in the stratopause region (for example, because only few data are available for the data assimilation), the missing drag can be used as a proxy for gravity wave drag, and its relative variations should still contain valuable information.

3 The SAO momentum budget in ERA-Interim

Altitude–time cross sections of the ERA-Interim zonal wind and the different terms of the tropical momentum budget in the altitude range 30–60 km are shown in Fig. 1. As mentioned before, all values in these time series are 7-day averages, additionally averaged over the latitude band 10° S–10° N. The time series covers the period from January 2005 until June 2008. Figure 1a shows the zonal-mean zonal wind, which displays a pronounced semiannual oscillation pattern, centered at about 47 km altitude. During the course of a calendar year, a strong westward wind phase is followed by an eastward wind phase, a weaker westward wind phase, and, again, an eastward wind phase. There is, however, considerable interannual variability in the strength of the different wind phases, as well as the exact timing and altitude of their maxima. A longer time series of tropical winds can be found, for example, in Dee et al. (2011), and an investigation of the relative

\[
\overline{X}_{PW} = \overline{X}_{\text{res}}(k \leq 20).
\]
strength of different SAO cycles has been carried out, for example, by Delisi and Dunkerton (1988). In particular, Delisi and Dunkerton (1988) argue that activity of Rossby waves at northern latitudes is responsible for the finding that the first SAO cycle of a year is often stronger. At altitudes below 40 km the zonal wind is dominated by the QBO.

Figure 1b shows the wind tendency $\frac{\partial u}{\partial t}$. Usually the zones of eastward (i.e., positive) wind tendency are tilted, meaning they descend in altitude with time. This characteristic behavior is typical for wind reversals that are driven by upward-propagating waves (e.g., Ern and Preusse, 2009a; Ern et al., 2013, 2014). Different from this, zones of westward (i.e., negative) wind shear are almost vertical. This indicates that SAO wind reversals from eastward to westward wind are likely not driven by dissipation of upward-propagating waves that have their sources in the tropical troposphere. In the period considered, there is, however, one exception to this rule: the westward shear zone in mid-2006 descends in altitude with time, which suggests that (as an exception) this wind reversal is mainly driven by upward-propagating tropical waves.

Even though fewer and fewer data are assimilated in ERA-Interim at increasing altitude, the SAO zonal winds should be quite reliable (e.g., Baldwin and Gray, 2005; Dee et al., 2011; Baron et al., 2013). Consequently, the zonal wind tendency $\frac{\partial u}{\partial t}$, which is determined directly from $\tilde{u}$, should also be quite reliable.

In ECMWF data planetary waves are quite realistic in the lower stratosphere (Timmermans et al., 2005; Feng et al., 2007; Ern et al., 2008; Yang et al., 2011). Although the quality of the planetary waves in ECMWF somewhat decreases toward higher altitudes (Ern et al., 2008, 2009b), it can be assumed that the main features of planetary wave driving are captured by ERA-Interim at stratopause heights.

ERA-Interim planetary wave drag for zonal wave numbers $k = 1–20$ derived from the EP flux divergence is shown in Fig. 1c. Satellite observations suggest that eastward-directed planetary wave drag of vertically propagating Kelvin waves (the strongest eastward-propagating equatorial planetary wave mode) should be small at the stratopause (e.g., Hitchman and Leovy, 1988; Ern and Preusse, 2009a). This is also indicated in the ERA-Interim planetary wave drag: eastward planetary wave drag is mostly weak. Except for sporadic events, it is usually weaker than 0.5 m s$^{-1}$ d$^{-1}$.

Westward planetary wave drag, however, is usually quite strong during early winter of both hemispheres (June/July and November/December), and it reaches values as high as $-2$ m s$^{-1}$ d$^{-1}$ (occasionally even $-4$ m s$^{-1}$ d$^{-1}$). These bursts of strong planetary wave drag are likely not caused by vertically propagating equatorially trapped waves, because these bursts do not show the characteristic descent in altitude with time that is typical for wave–mean-flow interaction by critical-level filtering of upward-propagating waves. Instead, the strong planetary wave drag events occur simultaneously over a large altitude range. This indicates that these events are likely caused by horizontal transport of wave momentum of extratropical waves from the polar jets, as has been proposed by several authors (e.g., Hirota, 1980; Delisi and Dunkerton, 1988). These strong bursts of planetary wave drag are in good correspondence with the periods of strong westward (negative) zonal wind tendency in Fig. 1b, and are therefore likely the main driver of the wind reversal from SAO eastward to SAO westward winds. One exception is mid-2006; during this period the westward-directed planetary wave drag is less pronounced.

The missing drag in the ERA-Interim momentum budget is displayed in Fig. 1d. This missing drag is the sum of wind tendency (Fig. 1b) and advection terms minus planetary wave drag (Fig. 1c). For a discussion of the ERA-Interim advection terms see Appendix A. Even if the magnitude of the missing drag may not be fully realistic, relative variations can provide some information about the contribution of gravity waves in the SAO momentum budget.

Planetary wave drag alone is almost sufficient to explain the negative (i.e., westward) wind tendencies in Fig. 1b. Since other contributions of negative drag are much weaker, the missing drag (Fig. 1d) is dominated by the meridional advection term, which is the strongest positive contribution in the ERA-Interim momentum budget (see Appendix A). Gravity waves are therefore expected to contribute mainly to the SAO wind reversals from westward to eastward winds. Westward (i.e., negative) gravity wave drag (i.e., missing drag) is usually much weaker and found only sometimes during westward wind shear (for example in mid-2006), or in the lower part of the SAO westward wind jets during December/January.

4 Gravity waves observed from satellite and the SAO momentum budget

We now investigate how observed gravity waves are modulated by the SAO, and whether observed absolute gravity wave momentum fluxes and gravity wave drag are in agreement with the theoretical picture of the driving of the SAO. In particular, it is expected that eastward wave driving of the SAO should be dominated by upward-propagating gravity waves, while westward driving is expected to arise from extratropical planetary waves (e.g., Reed, 1966; Hirota, 1980; Delisi and Dunkerton, 1988; Hamilton and Mahlmann, 1988).

4.1 Gravity wave squared amplitudes

Figure 2a shows an altitude–time cross section of gravity wave squared temperature amplitudes of the strongest gravity waves found in the individual altitude profiles in the latitude band $10^\circ$ S–$10^\circ$ N using a MEM/HA vertical analysis with a 10 km vertical window (see Sect. 2.1.2). Squared amplitudes were divided by 2 to make the values directly comparable to gravity wave temperature variances. For a compar-
Figure 2. Altitude–time cross sections of averages over the latitude band 10° S–10° N. (a) HIRDLS gravity wave squared temperature amplitudes determined in 10 km vertical windows from the HIRDLS altitude profiles. Values were divided by 2 to be comparable to gravity wave variances. Units in (a) are dB(K²). (b) HIRDLS total gravity wave momentum fluxes in mPa from a gravity wave analysis using a 10 km vertical window covering vertical wavelengths < 25 km, and (e) total gravity wave drag obtained from vertical gradients of the HIRDLS momentum fluxes shown in (b). For comparison, ERA-Interim (d) ∂u/∂t and (e) planetary wave drag (k = 1–20) are repeated from Fig. 1b and c, respectively. Units in (c–e) are m s⁻¹ d⁻¹. Contour lines indicate the zonal wind: westward wind is dashed, and the bold contour line indicates zero wind. Contour increment is 20 m s⁻¹.

Figure 3. Time series at 47 km (about the center altitude of the SAO) for the period January 2005 until June 2008. All parameters are averages over the latitude band 10° S–10° N. (a) The zonal-average zonal wind from ERA-Interim in m s⁻¹ and (b) the ERA-Interim zonal momentum budget terms: ∂u/∂t (black dashed), 0.5 times the sum of meridional and vertical advection terms (blue), 0.5 times missing drag (red), and planetary wave drag for zonal wave numbers 1–20 (green). (c) Comparison of observed HIRDLS gravity wave drag at 46 km (black) and absolute values of several ERA-Interim terms at 47 km averaged vertically over 10 km: ∂u/∂t (black dashed), planetary wave drag (green), and 0.5 times missing drag (red). (d) Range of ground-based phase speeds (red shaded) that are filtered out by the QBO in the altitude range 18–40 km. (e) ERA-Interim zonal wind at the altitude levels z₁ = 41 km (black) and z₂ = 51 km (red). Months of four characteristic background wind situations are indicated by blue hatched bands. In all panels periods of strong eastward (eastward) wind shear are indicated by gray (orange) shading. Times when the zonal wind at 47 km is zero are marked by brown vertical lines.
able propagation conditions (increased critical amplitudes because background wind and phase speed of the waves are opposite). Consequently, we find large variances and amplitudes of likely eastward-propagating gravity waves during SAO westward wind phases, much higher than gravity wave variances and amplitudes during SAO eastward winds.

4.2 Gravity wave momentum flux

The altitude–time distribution of absolute values of HIRDLS gravity wave momentum fluxes estimated as described in Sect. 2.1.2 is displayed in Fig. 2b. The values are given in millipascal on a linear scale. As already indicated by the enhanced gravity wave variances during phases of SAO westward winds, gravity momentum fluxes are also much stronger during westward winds than during eastward winds. Again, this is an effect of the filtering of the spectrum of upward-propagating gravity waves by the QBO in the stratosphere.

In the absence of wave dissipation, gravity wave momentum flux would be a conserved quantity. Different from this, in Fig. 2b momentum flux decreases continuously with altitude, indicating that there is always some dissipation of gravity waves at almost all altitudes and during most of the time. Another important finding is that the momentum flux maxima during SAO westward wind phases have a characteristic triangular (sawtooth-like) shape: the shape of these maxima follows the downward propagation of the zones of strong eastward wind tendencies with time. This indicates that the gravity waves dissipate and interact with the background winds. Similar effects have been observed before for Kelvin wave momentum fluxes during QBO eastward wind shear (e.g., Ern and Preusse, 2009a), for gravity wave momentum fluxes during both eastward and westward wind shear of the QBO (Ern et al., 2014), and for the wind reversal from mesospheric westward to eastward winds in the summer hemisphere at midlatitudes (Ern et al., 2013).

4.3 Gravity wave drag

Figure 2c shows altitude–time cross sections of absolute (total) gravity wave drag calculated from vertical gradients of absolute momentum fluxes. Around the stratopause gravity wave drag varies between about zero and somewhat above 1 m s\(^{-1}\) d\(^{-1}\). As expected, gravity wave drag usually maximizes during eastward-directed (i.e., positive) vertical shear of the zonal wind. This is particularly the case during December, January and February in each year, i.e., when the eastward shear is strongest.

As mentioned above, the spectrum of upward-propagating gravity waves has been filtered by the QBO before reaching the stratopause region, and usually westward-propagating gravity waves will undergo stronger filtering. Particularly for SAO westward wind phases it can therefore be assumed that the gravity wave distribution is dominated by waves propagating eastward, i.e., opposite to the SAO background wind. During eastward wind shear the propagation conditions of eastward-propagating waves become worse, and they will undergo stronger dissipation. Therefore the direction of gravity wave drag during eastward wind shear should be eastward.

This means, there is clear observational evidence that upward-propagating gravity waves contribute strongly to the reversal from SAO westward to SAO eastward winds. This also agrees well with the fact that the zones of eastward wind shear propagate downward with time.

During westward (i.e., negative) vertical shear of the zonal wind gravity wave drag is usually much weaker. Given the fact that eastward-propagating gravity waves should dominate the gravity wave momentum flux spectrum in the stratopause region, it is difficult to tell whether the gravity wave drag during westward shear is directed eastward or westward. Only in May and June 2006, at altitudes above 50 km, an enhancement of observed gravity wave drag follows closely the negative vertical shear of the zonal wind. Together with the fact that ERA-Interim missing (i.e., gravity wave) drag in this region is negative (see also Fig. 1d), this indicates that the observed gravity wave drag should also be negative (westward). Moreover, the exceptional descent of the zone of westward wind shear with time in mid-2006 seems to be mainly driven by dissipation of westward-propagating gravity waves.

For comparison with the satellite observations, Fig. 2d shows the zonal wind tendency \(\partial \bar{u} / \partial t\) in ERA-Interim, and Fig. 2f shows the planetary wave drag in ERA-Interim calculated from the EP flux divergence of resolved waves with zonal wave numbers \(1 \text{–} 20\).

By comparing Fig. 2c–e we find that negative (i.e., westward) values of ERA-Interim zonal wind tendency match very well with ERA-Interim planetary wave drag. Both the tendency and planetary wave drag show enhanced values in very short bursts that cover larger altitude regions in nearly vertical bands. At the same time, observed absolute gravity wave drag is usually small.

During periods when the zonal wind tendency is positive (i.e., directed eastward), however, the situation is reversed: the tendency shows maxima that descend in altitude with time. These maxima coincide with enhanced observed absolute gravity wave drag, while planetary wave drag is weak at the same time.

This indicates that SAO wind reversals from eastward to westward wind (westward tendency) are mainly driven by planetary waves (likely of extratropical origin), while the wind reversals from SAO westward wind to SAO eastward wind (eastward tendency) are mainly driven by eastward-propagating gravity waves of tropical origin.

However, there are also exceptions. For example, the eastward to westward wind reversal in mid-2006 seems to be mainly driven by eastward-propagating gravity waves. Further, negative values of gravity wave drag in ERA-Interim
(see Fig. 1d) might indicate that the dissipation of westward-propagating gravity waves could also be important on other occasions (for example, in the lower part of the SAO westward jet in December–February).

It should also be noted that the increases in observed gravity wave drag usually coincide with the ERA-Interim zonal wind shear zones. This means that the observed gravity wave drag is in reasonable agreement with the pattern that would be expected from wind filtering of the wave spectrum by the ERA-Interim SAO winds. In addition to previous findings (e.g., Baldwin and Gray, 2005; Dee et al., 2011; Baron et al., 2013), this is another indication that the SAO winds in ERA-Interim are quite realistic, at least in their basic features.

4.4 Time series in the stratopause region

The timing of the different terms in the SAO momentum budget is investigated in more detail for an altitude of 47 km, i.e., about the center altitude of the SAO. Figure 3a shows the zonal-mean wind at this altitude. The characteristic SAO pattern of alternating eastward and westward winds can be clearly identified with the stronger westward phase (i.e., momentum budget. The black-dotted curve is the tendency zero wind are marked by brown vertical lines.

Figure 3b shows the different terms in the ERA-Interim momentum budget. The black-dotted curve is the tendency of the zonal wind ($\partial u/\partial t$), the green line is the drag due to resolved planetary waves with zonal wave numbers 1–20, the blue line is the sum of the meridional and the vertical advection terms, and the red line is the gravity wave drag (missing drag) in the ERA-Interim momentum budget. Because both advection term and gravity wave drag are much stronger, they have been multiplied by 0.5 to make them better comparable to the other terms. Please note that the advection term and gravity wave drag are on different sides of Eq. (3) and partly compensate for each other.

Again, we find that most of the eastward wind tendency in ERA-Interim can be explained by gravity wave drag, and most of the westward wind tendency by planetary waves. Sometimes strong westward wind tendency results in strong negative peaks in the time series. These peaks often coincide with strong bursts of negative (i.e., westward-directed) planetary wave drag. Negative gravity wave drag during westward wind tendencies is usually weak. Only in mid-2006, and in the period December 2006 until January 2007, are there also stronger events of negative (i.e., westward) gravity wave drag. Other instances of stronger negative gravity wave drag seen in Fig. 1d, for example at the beginning of the years 2005, 2006 and 2008, are at lower altitudes and therefore do not show up in the time series at 47 km altitude.

In Fig. 3c HIRDLS observations are compared with ERA-Interim. In order to account for a minor observational filter effect, we compare the HIRDLS time series from 46 km with the ERA-Interim time series at 47 km (see also Ern et al., 2014). In Fig. 3c the black solid line shows the total gravity wave drag at 46 km derived from HIRDLS observations. Absolute values of the following ERA-Interim momentum terms at 47 km altitude are also displayed: zonal wind tendency (black dotted), planetary wave drag (green), and missing (gravity wave) drag (red). For better comparison, the ERA-Interim terms were smoothed vertically by a 10 km running mean and averaged over the latitude band 10° S–10° N after taking the absolute values. The vertical smoothing is applied to account for the 10 km vertical window of our HIRDLS gravity wave analysis. Again, gravity wave drag from ERA-Interim is multiplied by 0.5 for better comparison.

Of course, by taking the absolute value and by averaging vertically over the different ERA-Interim terms, significant information is lost. Still, in Fig. 3c, there is an overall correspondence between enhancements of absolute zonal wind tendency and absolute planetary wave drag during periods of negative wind tendency, as well as between the relative variations in absolute zonal wind tendency and absolute ERA-Interim gravity wave drag during periods of positive wind tendency. There is also reasonable correspondence between the relative variations in ERA-Interim and HIRDLS absolute gravity wave drag. The correlation coefficient between absolute ERA-Interim and HIRDLS gravity wave drag is 0.77, which is highly significant, given the high number of data points (around 380; effectively only around 160 due to overlapping bins).

In addition, for both HIRDLS and ERA-Interim the integral over the gravity wave drag peak in the beginning of each year is larger than the integral over the peak in the middle of each year. This is qualitatively in good agreement with the fact that usually the westward wind phase in the beginning of each year is stronger, and more gravity wave drag is required for the wind reversal to eastward wind. This indicates that relative variations in ERA-Interim gravity wave drag might still contain useful information at 47 km altitude.

Absolute values of ERA-Interim gravity wave drag, however, are usually much higher: at least a factor of 2, and sometimes a factor of 4. Even though HIRDLS observes only part of the whole spectrum of gravity waves (only horizontal wavelengths > 100–200 km), this difference is probably too high to be explained by observational filter effects alone. This is further supported by several modeling studies that obtain much weaker gravity wave drag than ERA-Interim (e.g., Scaife et al., 2002; Osprey et al., 2010; Peña-Ortiz et al., 2010).

In Fig. 3d the range of zonal wind speeds in the altitude range 18–40 km is indicated by an area shaded in red. This is about the range of ground-based zonal phase speeds that should be removed from the spectrum of all gravity waves via critical-level filtering by the QBO winds. Because the westward wind phase of the QBO is much stronger, usually
westward phase speeds as strong as \(-40 \text{ m s}^{-1}\) are removed from the spectrum. On the other hand, almost all eastward-propagating gravity waves with phase speeds higher than about \(10 \text{ m s}^{-1}\) should still be contained in the gravity wave spectrum entering the altitude range where the SAO is observed.

One exception is the period April until July 2006. In this period only westward-propagating gravity waves with phase speeds that are less negative than \(-10 \text{ m s}^{-1}\) are filtered out by the QBO. It can therefore be expected that more westward-directed gravity wave momentum flux is available for the driving of the SAO than during other periods. In addition, during this period the drag due to planetary waves is comparably weak (see Fig. 1c). This probably explains why in mid-2006 exceptionally a downward-propagating westward shear zone develops. This shear zone is likely driven by westward-propagating gravity waves originating from the tropics, and not by planetary waves from the extratropics like the other westward wind reversals in ERA-Interim in the period 2005 until mid-2008. This expected enhancement of westward gravity wave drag is clearly seen in the HIRDLS absolute gravity wave drag shown in Fig. 2c; however it is only weakly indicated in Fig. 3c because the westward gravity wave drag maximizes at altitudes higher than 46 km. Possibly, the stronger than usual westward gravity wave drag also contributes to the fact that westward SAO winds in mid-2006 are somewhat stronger than in the middle of the other years in the period considered in our study.

Of course, it should be emphasized that the situation in mid-2006 is an exceptional event. Even if a longer period of over 10 years (2002–2012) is considered, this is the only event of this strength. This indicates that such events are likely not important from a climatological point of view. Still, from this event we can learn more about the effect of the QBO on the driving of the SAO.

### 5 Discussion of four characteristic cases

In the following, we want to obtain a better qualitative understanding of the dissipation of gravity waves in the stratosphere region. With this improved physical understanding, we will be able to identify the most relevant processes that should be included in global models for simulating a realistic SAO. Therefore, in this section, we will qualitatively discuss characteristic situations of the atmospheric background and the gravity wave dissipation that may result from this. Some evidence for this understanding will be presented in Sect. 6 by discussing observed gravity wave spectra. During all of these considerations we will focus on the zonal direction only (zonal winds and zonally propagating gravity waves), because meridionally propagating gravity waves will not contribute much to the SAO.

Figure 3e shows the SAO winds at 41 km (“level 1”, black curve) and 51 km altitude (“level 2”, red curve). There are four basic cases that can be identified:

- a. negative (i.e., westward) vertical shear of the zonal wind between levels 1 and 2 – zonal wind at level 1 is negative (i.e., westward);
- b. positive (i.e., eastward) vertical shear of the zonal wind between levels 1 and 2 – zonal wind at level 1 is negative (i.e., westward);
- c. negative (i.e., westward) vertical shear of the zonal wind between levels 1 and 2 – zonal wind at level 1 is positive (i.e., eastward);
- d. positive (i.e., eastward) vertical shear of the zonal wind between levels 1 and 2 – zonal wind at level 1 is positive (i.e., eastward).

Here, “positive” and “negative” shear refer to the average vertical shear considering the whole altitude range between level 1 and level 2. Four months approximately matching these characteristic cases are indicated in Fig. 3e by blue hatched rectangles: (a) January 2006, (b) March 2006, (c) June 2006, and (d) May 2007. Zonal wind altitude profiles for these months are shown later in Fig. 5 (see Sect. 6). However, before addressing these real-world situations, we will qualitatively discuss “idealized” cases.

The “idealized” situation of these four cases is illustrated in Fig. 4 with a schematic picture for each case. In each of Fig. 4a–d, the x axis indicates the zonal wind speed \(u\) and the gravity wave ground-based phase speed \(c_y\). The y axis is the vertical coordinate \(z\), and it also stands for the strength of gravity wave momentum flux, indicated by the vertical extent of the blue hatched rectangles. Two altitude levels are highlighted as \(z_1\) and \(z_2\). Level \(z_1\) is assumed to be situated directly on top of the region dominated by the QBO, while \(z_2\) is assumed to be situated in the altitude region dominated by the SAO. (In our work we assume the levels \(z_1 = 41\) km and \(z_2 = 51\) km.) The zonal wind vertical profile between the levels \(z_1\) and \(z_2\) is indicated by a red line. For simplification, it is assumed that the zonal wind changes monotonously with altitude, i.e., has a constant vertical gradient. The range of gravity wave phase speeds that is assumed to be filtered out by the QBO at altitudes \(z < z_1\) is marked by two vertical green dashed lines.

The amount of eastward-directed (MFe) and westward-directed (MFw) gravity wave momentum fluxes on the levels \(z_1\) and \(z_2\) is qualitatively indicated by blue hatched rectangles. The extent of the rectangles in the x direction gives the range of gravity wave phase speeds, while the extent in the y direction is a measure for the amount of momentum flux.
at a given phase speed. Figure 4a and b are for westward-directed zonal wind at the lower altitude level \( u(z_1) < 0 \), resulting in enhanced momentum flux \( MFe(z_1) \) and reduced \( MFw(z_1) \). Different from this, Fig. 4c and d represent cases of \( u(z_1) > 0 \), resulting in reduced \( MFe(z_1) \) and enhanced \( MFw(z_1) \). Consequently, this dependency on the background wind means that the direction of QBO winds (i.e., the QBO phase) at \( z = z_1 \) has a strong influence on the amount of eastward and westward-directed momentum fluxes at this altitude.

One of the limiting factors of momentum flux is wave saturation. Due to the decrease in air density, the amplitude of a conservatively propagating gravity wave grows exponentially with altitude. At some point, however, the amplitude cannot grow further and reaches its saturation limit, and thereafter the wave starts to dissipate. The saturation amplitude \( \hat{T}_{sat} \) is proportional to the difference between ground-based phase speed and background wind. In the following, we only consider the zonal direction, because this the only direction that is relevant for the driving of the SAO. In this case, the saturation amplitude is given by

\[
\hat{T}_{sat} = \frac{T}{g} N |c_\psi - \bar{u}|. \tag{10}
\]

See also Eq. (10) in Ern et al. (2008).

Because the temperature amplitude enters Eq. (1) in a quadratic way, \( \hat{T}_{sat} \) is also limiting the momentum flux of a gravity wave. Overall, the momentum flux of a saturated gravity wave is proportional to the third power of \( |c_\psi - \bar{u}| \) (see Eqs. 1 and 4 in Preusse et al., 2006), which shows the importance of the background winds. For the special case of a critical wind level \( (c_\psi = \bar{u}) \), the critical amplitude becomes zero, and the wave dissipates completely. However, saturation and wave dissipation can also occur without critical wind levels being reached. For a review on saturation effects of gravity waves see, for example, Fritts (1984).

Considering the whole spectrum of gravity waves in a given propagation direction (in our case either eastward or

---

Figure 4. Schematic illustration of four characteristic cases how gravity wave drag contributions in the SAO momentum budget are generated. Two altitude levels are considered. The level \( z_1 \) is located below the SAO-related wind shear, and the level \( z_2 \) above it. The red lines indicate simplified vertical profiles of the zonal background wind \( u \). The blue hatched boxes indicate the amount of gravity wave momentum flux at the two levels with \( MFw \) (\( MFe \)) the momentum flux for westward (eastward) ground-based phase speed \( C_\psi \). It is assumed that there is no momentum flux at ground-based phase speeds located between the green dashed vertical lines, because this phase speed range has been removed through critical-level filtering by the QBO at altitudes \( z < z_1 \). The four cases are (a) westward vertical shear of the zonal wind, \( u(z_1) < 0 \) and \( u(z_2) < 0 \); (b) eastward wind shear, \( u(z_1) < 0 \) and \( u(z_2) > 0 \); (c) westward wind shear, \( u(z_1) > 0 \) and \( u(z_2) < 0 \); and (d) eastward wind shear, \( u(z_1) > 0 \) and \( u(z_2) > 0 \).
westward), the gravity waves propagating into this direction can attain larger amplitudes and thereby carry more momentum if the intrinsic phase speed $|c_{\phi} - \langle u \rangle|$ and thus the saturation amplitude is high.

Therefore, for the scheme in Fig. 4, we assume that, at a given altitude level, the overall momentum flux for a given propagation direction can be higher if the difference $|c_{\phi} - \langle u \rangle|$ is high. This is particularly the case when background wind and ground-based phase speed have opposite directions.

Further, gravity wave observations show that gravity wave momentum fluxes continuously decrease with altitude (e.g., Ern et al., 2011). Therefore, we assume in Fig. 4 that gravity wave momentum flux at the higher altitude level $z_2$ should always be lower than at the lower altitude level $z_1$.

Of course, the use of rectangles for the shape of the gravity wave spectrum is just an illustration, and the “true” gravity wave spectrum will have a much more complicated shape, depending on the details of the gravity wave sources and the gravity wave dissipation at altitudes below $z_1$. For example, the study by Beres et al. (2005) indicates that there could be an asymmetry of the gravity wave spectrum in the tropopause region, with much stronger momentum fluxes at eastward-directed gravity wave phase speeds. If this is the case, it would be expected that, on average, the forcing in the stratopause region should be directed even more prevalently eastward than already expected from the stronger filtering of westward-propagating gravity waves by the QBO. Consideration of such effects is, however, beyond the scope of our study. The very simplified scheme proposed here is only intended to explain the very basic mechanisms leading to the observed effects. In particular, there is still considerable uncertainty about the shape of the momentum flux spectrum of convective gravity waves (see also Choi et al., 2012; Ern and Preusse, 2012).

Based on our very simplified assumptions, implications for gravity wave drag will now be discussed separately for each of the four cases introduced at the beginning of this section.

5.1 Case (a), January 2006: $u(z_1) < 0$, $\partial u/\partial z < 0$

At the level $z_1$ the momentum flux $M_{Fe}(z_1)$ of gravity waves with eastward-directed phase speed is high because $|c_{\phi} - \langle u \rangle|$ is high. At the same time the momentum flux $M_{Fw}(z_1)$ of gravity waves with westward-directed phase speed is comparably low because $|c_{\phi} - \langle u \rangle|$ is low (see Fig. 4a).

With increasing altitude, propagation conditions for gravity waves with eastward-directed phase speeds become more favorable because, for those waves, $|c_{\phi} - \langle u \rangle|$ and thus $T_{sat}$ increases with altitude. Because the waves already have large amplitudes at $z = z_1$, it is nevertheless expected that, at some point, a considerable part of them will reach their saturation amplitude and start to dissipate, even though their intrinsic phase speeds are high and they do not encounter critical wind levels. Due to the fact that $M_{Fe}(z_1) \gg M_{Fw}(z_1)$, the resulting net gravity wave drag will be strongly positive (i.e., eastward).

This wave saturation effect should happen preferentially at high altitudes (close to $z_2$) because it depends on the amplitude growth of the gravity waves with altitude. Indeed, such strong values of gravity wave drag are seen in Fig. 2c in HIRLDS observations at altitudes above about 45 km during January 2006 and other situations matching case (a). In these cases we also find strongly positive values of ERA-Interim missing drag (see Fig. 1d).

Still, from Fig. 1d there are also indications for weak negative (i.e., westward-directed) gravity wave drag at low altitudes (between about 40 and 45 km) during periods matching the conditions of case (a), for example during January 2006 and January 2007. This finding could be realistic, because in case (a), for gravity waves with westward-directed phase speeds, the difference $|c_{\phi} - \langle u \rangle|$ reduces with altitude, leading to lowered saturation amplitudes and enhanced dissipation. Part of these gravity waves will even encounter critical levels where $|c_{\phi} - \langle u \rangle| = 0$. Although $M_{Fe}(z_1) \gg M_{Fw}(z_1)$, at low altitudes the dissipation of westward-propagating gravity waves might still dominate and lead to slightly negative gravity wave drag.

Overall, there is strong indication that the driving of the SAO cannot be understood alone from critical-level filtering of gravity waves between the two levels $z_1$ and $z_2$. It is very likely that gravity wave saturation without critical levels being reached also plays an important role. Otherwise the strong values of eastward gravity wave drag that always occur in case (a) at high altitudes (close to $z_2$) cannot be ex-
explained. This will be further discussed in Sect. 6.1, and more observational evidence will be presented.

5.2 Case (b), March 2006: $u(z_1) < 0$, $\partial u/\partial z > 0$

Like in case (a), the momentum flux $MFe(z_1)$ is high and $MFw(z_1)$ is low (see Fig. 4b). With increasing altitude the propagation conditions become worse (better) for gravity waves with eastward (westward-) directed phase speeds as $|c_p - \bar{u}|$ decreases (increases). Since at $z = z_1$ $MFe(z_1) \gg MFw(z_1)$, the resulting net gravity wave drag should therefore clearly be positive (directed eastward).

5.3 Case (c), June 2006: $u(z_1) > 0$, $\partial u/\partial z < 0$

In case (c) the momentum flux $MFw(z_1)$ is high, and $MFe(z_1)$ is low, because $|c_p - \bar{u}|$ is high for westward-propagating gravity waves, and low for eastward-propagating gravity waves (see Fig. 4c). With increasing altitude the propagation conditions become worse (better) for gravity waves with westward- (eastward-) directed phase speeds as $|c_p - \bar{u}|$ decreases (increases). Although the phase speed spectrum of westward gravity waves is more strongly filtered by the QBO, and only waves with high ground-based phase speeds are remaining, the resulting net gravity wave drag should be negative (directed westward).

Because usually westward-directed momentum fluxes are partly filtered out by the QBO, it would be expected that the resulting net gravity wave drag is weaker than, for example, in case (a). This is also indicated in Fig. 3c for a fixed altitude of 47 km.

5.4 Case (d), May 2007: $u(z_1) > 0$, $\partial u/\partial z > 0$

At the level $z_1$ the momentum flux $MFe(z_1)$ of gravity waves with eastward-directed phase speed is low because $|c_p - \bar{u}|$ is low. At the same time, the momentum flux $MFw(z_1)$ of gravity waves with westward-directed phase speed is high because $|c_p - \bar{u}|$ is high (see Fig. 4d). However, the phase speed range of $MFw$ is strongly reduced due to filtering of the gravity wave spectrum by the QBO at altitudes $z < z_1$. With increasing altitude the difference $|c_p - \bar{u}|$ is increased for gravity waves with westward-directed phase speeds, resulting in only little westward-directed gravity wave drag. At the same time, $|c_p - \bar{u}|$ is reduced for gravity waves with eastward-directed phase speeds. Nevertheless, this results in only little gravity wave drag, because $MFe(z_1)$ is already low at $z = z_1$. In particular, during the period considered in our study, eastward wind phases of the SAO are usually weaker than westward wind phases. Therefore, the vertical gradient $\partial u/\partial z$, and also the resulting (net) gravity wave drag, will only be weak in case (d).

6 Gravity wave spectra in the four characteristic cases

We now discuss gravity wave momentum flux spectra observed by HIRDLS for conditions roughly corresponding to the four cases introduced in Sect. 5. In particular, an interesting question is whether there is any evidence of gravity wave saturation effects without critical levels being reached. If this is the case, this might have important implications for the representation of gravity waves in global models, either resolved or parameterized.

As already mentioned, the situations of cases (a)–(d) are roughly matched in (a) January 2006, (b) March 2006, (c) June 2006, and (d) May 2007, respectively. Vertical profiles of the zonal wind for these months are shown in Fig. 5 (blue curves). Of course, the zonal wind vertical profiles in Fig. 5 only on average match the idealized assumption made in Sect. 5 of a linear increase or decrease in the zonal wind with altitude. Nevertheless, the change in the background winds on average, and the corresponding change in observed gravity wave momentum flux spectra, will provide further insight into details of the forcing of the SAO.

From single observations of gravity wave momentum flux, the average gravity wave spectrum in a certain region can be recovered (see also Lehmann et al., 2012). For example, momentum flux as function of horizontal and vertical wave number can be determined by sorting the single observed momentum fluxes into bins in the plane of horizontal and vertical wave numbers (see also Ern and Preusse, 2012; Ern et al., 2014).

The determination of the HIRDLS gravity wave momentum flux spectra shown in Fig. 6 is very similar to that in Ern et al. (2014). The different rows in Fig. 6 correspond to the different cases described in Sect. 5. Figure 6a, d, g, and j (left column in Fig. 6) show momentum flux spectra at the altitude $z_1 = 41$ km, and Fig. 6b, e, h, and k (middle column in Fig. 6) show spectra at $z_2 = 51$ km. For both the left and the middle column, a logarithmic momentum flux scale is used. The right column (Fig. 6c, f, i, and l) shows, on a linear momentum flux scale, the difference between the spectra in the left and middle column. The spectra in the right column therefore provide information about the part of the wave spectrum that has dissipated between the altitudes $z_1$ and $z_2$. Of course, our analysis uses a 10 km vertical window, and all spectra shown in Fig. 6 represent only average conditions for altitude ranges of 10 km. Therefore, they will only on average match the four idealized cases. The vertical intervals and average zonal winds that correspond to the spectra in Fig. 6 are marked in Fig. 5 by red vertical bars that are centered at the altitude levels $z_1 = 41$ km and $z_2 = 51$ km, respectively.

6.1 Case (a), January 2006: $u(z_1) < 0$, $\partial u/\partial z < 0$

The situation of case (a) is approximately matched during January 2006. The corresponding spectra are shown in Fig. 6a–c. The reduction of momentum fluxes shown
While the reduction of momentum fluxes at low phase speeds at intrinsic phase speeds westward phase speeds, significant reduction of momentum fluxes is also found at vertical wavelengths $> 10$ km, i.e., intrinsic phase speeds $> 60$ m s$^{-1}$. The latter cannot be explained by critical-level filtering of westward phase speeds. Obviously gravity waves with high intrinsic eastward phase speeds also dissipate and possibly reach their saturation amplitude.

This is further evidence that, in case (a) close to the level $z_2$ in the upper part of the SAO region, gravity wave drag should be strongly positive (i.e., eastward) (see also Fig. 1d). In addition, this indicates that, even though the vertical gradient of the zonal wind between the levels $z_1$ and $z_2$ is strongly negative on average, eastward-propagating gravity waves reach saturation (because these waves have quite high amplitudes and strong momentum fluxes). This will be even more likely the case for situations like in Fig. 5a when the local vertical gradient of the zonal wind weakens or reverses close to $z_2$. While saturation and dissipation of eastward-propagating gravity waves will strongly dominate at high altitudes, at low levels (close to $z_1$) dissipation of westward-propagating gravity waves could still result in slightly negative (i.e., westward-directed) net gravity wave drag, which is also indicated in Fig. 1d. Further, the involvement of such high intrinsic phase speeds shows that the background winds should be quite strong. This indicates that the quite strong ERA-Interim winds during January 2006 (stronger than $-60$ m s$^{-1}$ at 47 km altitude) might be realistic.

6.2 Case (b), March 2006: $u(z_1) < 0$, $\partial u/\partial z > 0$

Background conditions matching case (b) are found during approximately March 2006, and the corresponding spectra are shown in Fig. 6d–f. Compared to case (a), the reduction of momentum fluxes (Fig. 6f) is now shifted toward lower intrinsic phase speeds. Not much reduction is found at vertical wavelengths $> 15$ km (intrinsic phase speeds $> 45$ m s$^{-1}$), and the peak reduction is at vertical wavelengths $< 10$ km, i.e., intrinsic phase speeds $< 30$ m s$^{-1}$. This is in good agreement with the assumption that mainly gravity waves with eastward phase speeds undergo critical-level filtering, and gravity waves with westward-directed phase speeds should not contribute much. Accordingly, the resulting net drag should be positive (i.e., eastward), which is in good agreement with ERA-Interim during March 2006 in the altitude range 40–50 km (see Fig. 1d).

6.3 Case (c), June 2006: $u(z_1) > 0$, $\partial u/\partial z < 0$

The situation of case (c) is approximately matched during June 2006. The corresponding spectra are displayed in Fig. 6g–i. In case (c), the spectral distribution of the momentum flux difference between lower and upper level (Fig. 6i) is qualitatively very similar to case (b) (Fig. 6f). Absolute values are, however, somewhat reduced. Accordingly, this indicates that mainly gravity waves with westward phase speeds should undergo critical-level filtering, and gravity waves with

Figure 6. HIRDLS gravity wave momentum flux spectra for the four cases illustrated in Fig. 4. Upper row (a–c): January 2006, westward vertical shear of the zonal wind, $u(z_1) < 0$ and $u(z_2) < 0$; second row (d–f): March 2006, eastward wind shear, $u(z_1) < 0$ and $u(z_2) > 0$; third row (g–i): June 2006, westward wind shear, $u(z_1) > 0$ and $u(z_2) < 0$; and lower row (j–l): May 2007, eastward wind shear, $u(z_1) > 0$ and $u(z_2) > 0$. The left column (a, d, g, j) shows spectra for 41 km altitude, and the middle column (b, e, h, k) for 51 km altitude. These altitudes correspond to the lower and upper levels $z_1$ and $z_2$, respectively, in Fig. 4. The right column (c, f, i, l) shows the differences between the spectra at 41 km and the spectra at 51 km altitude.
eastward-directed phase speeds should not contribute much. In particular, Fig. 6i does not indicate a strong reduction of momentum fluxes at long vertical wavelengths (i.e., high intrinsic phase speeds), which would be an indication of saturation and dissipation of eastward-propagating gravity waves like in Fig. 6c (case (a)). The resulting net drag should therefore be negative (i.e., westward), which is also indicated in ERA-Interim during June 2006 in the altitude range 45–50 km (see Fig. 1d). In the observations peak values of drag during June 2006 are at somewhat higher altitudes (between 50 and 55 km; see Fig. 2c). Although momentum flux differences in Fig. 6i are somewhat reduced compared to Fig. 6c and f, still considerable drag is seen during June 2006 in Fig. 2c. The likely reason for this is the shift of peak drag toward higher altitudes: due to the decreased atmospheric density at higher altitudes, even a smaller amount of momentum flux can produce significant drag.

6.4 Case (d), May 2007: \( u(z_1) > 0, \partial u/\partial z > 0 \)

Conditions matching case (d) are found, for example, during May 2007. The corresponding spectra are shown in Fig. 6j–l. The reduction of momentum fluxes (Fig. 6i) is very weak compared to all other cases. Accordingly, only little gravity wave drag is observed in Fig. 2c. Further, ERA-Interim shows only weak drag during May 2007 (see Fig. 1d).

7 Conclusions

In our study we have investigated the momentum budget of the semiannual oscillation of the zonal wind (SAO) in the tropical latitude band 10°S–10°N for the period January 2005 until mid-2008. The main focus was on the contribution of gravity waves that is not easily accessible and is subject to large uncertainties in both observations and modeling studies.

Temperature observations of the High Resolution Dynamics Limb Sounder (HIRDLS) satellite instrument were used to derive absolute values of gravity wave momentum flux, as well as total (absolute) gravity wave drag from momentum flux vertical gradients. These values of gravity wave drag were compared to the different terms in the momentum budget of the ERA-Interim reanalysis of the European Centre for Medium-Range Weather Forecasts (ECMWF), in particular to the zonal wind tendency and the missing drag. It is assumed that this missing drag can be attributed to the gravity wave contribution in the momentum budget.

Based on our measurements and gravity wave theory we infer a consistent picture of the SAO. Our findings confirm the general assumption that gravity waves should mainly contribute to the SAO momentum budget during eastward wind shear. This is compatible with the filtering of the spectrum of upward-propagating waves by the QBO in the lower and middle stratosphere (e.g., Hirota, 1980; Delisi and Dunkerton, 1988; Hamilton and Mahlmann, 1988; Garcia et al., 1997). Background winds during QBO westward phases are usually much stronger than during eastward phases, and gravity waves with ground-based phase speeds between about −40 and 10 m s\(^{-1}\) tend to be filtered out by QBO wind before entering the upper stratosphere. This means that a large part of the spectrum of waves with westward-directed (i.e., negative) phase speeds is filtered out. Accordingly, we find that observed gravity wave variances and momentum fluxes are much stronger during SAO westward winds, which provide favorable propagation conditions for gravity waves with eastward-directed phase speeds. These waves are much less affected by the QBO than gravity waves with westward-directed phase speeds. The dissipation of gravity waves with eastward-directed phase speeds strongly contributes to the SAO momentum budget during SAO eastward wind shear, and in ERA-Interim, zones of SAO eastward wind shear propagate downward with time, as would be expected for wave-driven wind reversals.

These situations of eastward wind shear (\( \partial u/\partial z > 0 \)) prevail during much of the time of significant SAO gravity wave forcing, and they are discussed in more detail with the introduction of the two characteristic cases (b) and (d) in Sect. 5. These two cases differ in their direction of the zonal wind at an altitude level of \( z = z_1 \) located directly below the altitude region dominated by the SAO. The zonal wind at this altitude has a strong influence on the amount of eastward-directed gravity wave momentum flux that is available for driving the SAO. In particular, the gravity wave driving of the SAO is much stronger if the zonal wind at \( z = z_1 \) is westward. Consequently, the direction and strength of the QBO winds (i.e., the QBO phase) at \( z = z_1 \) plays an important role in the driving of the SAO.

During eastward wind shear, peak values of observed HIRDLS gravity wave drag are about 1 m s\(^{-1}\) d\(^{-1}\). This is qualitatively in good agreement with several modeling studies (e.g., Scaife et al., 2002; Osprey et al., 2010; Peña-Ortiz et al., 2010). Observed values are, however, somewhat lower, which may be explained by the fact that the observed drag represents an average over an altitude range of 10 km. In addition, our method may underestimate momentum fluxes. Further, HIRDLS observes only part of the whole spectrum of gravity waves, and therefore only part of the gravity wave drag. The missing drag in ERA-Interim is significantly higher than in the observations and in the other modeling studies. A likely reason for this is that the meridional advection is somewhat too strong in ERA-Interim. Nevertheless, relative variations in the missing drag in ERA-Interim provide some information, and there is good agreement between relative variations in observed gravity wave drag and relative variations in absolute values of ERA-Interim missing drag.

During westward wind shear, gravity wave drag is usually weaker in both observations and ERA-Interim, and in the ERA-Interim momentum budget the westward-directed (i.e., negative) zonal wind tendency is mainly balanced by plane-
tary wave drag. This planetary wave drag is likely of extratropical origin because zones of westward-directed SAO wind tendency are almost vertical and do not gradually propagate downward with time, as would be expected for wind reversals that are mainly driven by purely vertically propagating waves.

Still, we find exceptions where gravity waves may exert westward-directed drag in the SAO momentum budget. One of these exceptions is found during a period of westward-directed wind shear ($\partial u/\partial z < 0$); this is discussed in our characteristic case (c): in a time window during May and June 2006, the QBO filtering of waves with westward-directed phase speeds is less effective, and the drag due to planetary waves is found to be comparably weak in ERA-Interim. During this period the ERA-Interim missing drag is negative (westward). At the same time, absolute gravity wave drag observed by HIRDLS is high, and the zone of westward-directed SAO wind shear exceptionally shows a downward propagation with time. Other cases of westward-directed gravity wave drag might be during December and January in several years at altitudes 40–45 km (in the lower part of the SAO altitude region), as indicated in the ERA-Interim missing drag. It is, however, not clear whether this finding is very reliable.

In Sect. 5, another characteristic case (case (a)) is introduced, which addresses the situation of westward-directed wind at $z = z_1$ and westward-directed wind shear ($\partial u/\partial z < 0$). In this case, likely wave saturation of eastward-propagating gravity waves results in strong eastward-directed gravity wave drag, although the overall wind shear is westward, considering a larger altitude range. Case (a) is somewhat different from our other characteristic cases (b)–(d), because the intrinsic phase speeds of the dissipating eastward-propagating waves are high and critical wind levels are not being reached. This situation often occurs during January and may be important for the onset of SAO wind reversals from westward to eastward winds at high altitudes.

It is also notable that the location and timing of the shear zones of the ERA-Interim zonal wind reasonably well match the enhancements of observed gravity wave drag. In addition to previous studies (e.g., Baldwin and Gray, 2005; Dee et al., 2011; Baron et al., 2013), this is another indication that the ERA-Interim zonal wind in the stratopause region should be realistic in its basic features. Of course, the fact that ERA-Interim winds in the tropics are quite reasonable is a merit of the data assimilation scheme. Even though the model physics in the stratopause region is oversimplified (use of Rayleigh friction and strong damping of resolved waves above about 40 km altitude instead of a dedicated gravity wave parameterization scheme) quite reliable winds are simulated in the tropical stratopause region.

Free-running global models, however, cannot benefit from data assimilation and require a realistic representation of the most relevant physical processes. As has become apparent during our study, critical-level filtering of gravity waves is not sufficient for simulating a realistic SAO. Additionally, wave saturation processes without critical levels being reached play an important role and have to be parameterized in a realistic way. This implies that a realistic wave saturation scheme is required for the gravity waves explicitly resolved in the model. There are even indications that, depending on the model setup, data assimilation of lower atmospheric data is not always able to overcompensate the effect of an unrealistic gravity wave parameterization (e.g., Pedatella et al., 2014), which underlines the importance of including realistic physical processes in the models.

Overall, our study for the first time provides direct observational evidence from global observations of gravity waves that, indeed, gravity waves contribute strongly to the eastward wind reversals of the SAO but only weakly to the westward wind reversals, as would be expected from theoretical considerations. However, there are also exceptions when westward-directed gravity wave drag is important. Obviously, the momentum budget of the SAO is somewhat more complicated than expected. The findings of our study therefore provide important information and can give some experimental guidance to model studies and simulations of the SAO.

Of course, one of the drawbacks of our study is that the direction of gravity wave momentum fluxes cannot be directly inferred from the satellite observations. This is the case because information is provided only for the vertical direction and the direction along the satellite measurement track, i.e., only 2-D information is available for investigating the observed gravity waves. Therefore, there is still some uncertainty about the direction and magnitude of net gravity wave drag in general. This limitation could be overcome, for example, by the infrared limb-imaging technique, giving full 3-D information about the observed gravity waves by additionally providing observations for the direction across the satellite measurement track. For a more detailed discussion of this measurement technique and its capabilities see, for example, Riese et al. (2005, 2014) or Preusse et al. (2009, 2014).

Another approach for improving the representation of the SAO in global models would be a more systematic monitoring of temperatures and winds in the stratopause region, or even the mesosphere. In particular, global observations of winds in the stratopause region are sparse (see also Baron et al., 2013). Including such observations in the assimilation schemes of operational meteorological analyses, or in reanalysies, would improve their winds in this altitude region. These improved winds, in turn, could then serve as a reference for free-running global models and help to improve model physics, resulting in an improved simulated SAO.
Appendix A: Discussion of the advection terms in the ERA-Interim momentum budget

Figure A1a shows the vertical advection term of ERA-Interim. The main contributions of vertical advection are directed westward (up to $-2 \text{ m s}^{-1} \text{ d}^{-1}$), and they occur usually in the period December/January. There is also another event of strong westward forcing in mid-2006. Eastward forcing is usually weak.

The meridional advection term is shown in Fig. A1b. The forcing due to meridional advection is directed eastward and mainly occurs in the westward wind phases of the SAO. In addition, meridional advection strongly increases at altitudes above about 55 km. In the altitude range considered in our study the contribution of meridional advection can be as strong as about $5 \text{ m s}^{-1} \text{ d}^{-1}$.

Compared to other model simulations of the SAO, the advection terms in ERA-Interim are quite strong. In other simulations typical values in the stratosphere region are of the order $2.5 \text{ m s}^{-1} \text{ d}^{-1}$ and less, i.e., considerably weaker (e.g., Scaife et al., 2002; Osprey et al., 2010; Peña-Ortiz et al., 2010). Therefore the advection terms in ERA-Interim might not be fully realistic. Still, the relative variations should provide some information about the momentum budget in the stratosphere region. At higher altitudes (above 55 km), however, meridional advection seems to be no longer reliable.

The missing drag in the ERA-Interim momentum budget is shown in Fig. A1c (see also Fig. 1d). Obviously, the advection terms are the main contributions in the ERA-Interim missing drag. This means that, like the advection terms, the missing drag will not be fully realistic; however it still may provide useful information from its relative variations.

Appendix B: Comparison of gravity wave variances and squared amplitudes

Figure B1a shows an altitude–time cross section of HIRDLS temperature variances due to gravity waves, directly after removal of the large-scale atmospheric background temperatures (the first step as described in Sect. 2.1.2). All HIRDLS altitude profiles in the latitude band $10^\circ \text{ S}–10^\circ \text{ N}$ are considered. The variances are given in dB($K^2$), i.e., on a logarithmic scale. For comparison, Fig. B1b replicates Fig. 2a and shows gravity wave squared amplitudes of the strongest gravity waves found in each altitude profile in the latitude band $10^\circ \text{ S}–10^\circ \text{ N}$ using a MEM/HA vertical analysis with a 10 km vertical window (see Sect. 2.1.2). Squared amplitudes were divided by 2 to make the values directly comparable to the variances shown in Fig. B1a. We find that the distributions of variances and squared amplitudes are very similar. Obviously, even though a 10 km vertical window was applied for the determination of wave amplitudes, the squared amplitudes capture the basic features of the interaction of the gravity wave distribution with the SAO winds. Values in Fig. B1b are only slightly lower than in Fig. B1a (about 1.5 dB, i.e., 30%; please note that the color scale in Fig. B1b has been shifted). One reason for the slightly reduced values is an averaging effect of the 10 km vertical window. In addition, part of the gravity wave variance is carried by weaker waves that are neglected in our study. For the calculation of gravity wave momentum fluxes, only pairs of altitude profiles with matching vertical wavelengths are considered (see also Sect. 2.1.2). The distribution for squared amplitudes times 0.5 of these pairs looks almost exactly the same as in Fig. B1b (both in...
absolute values and relative variations) and is therefore not shown. The good agreement between gravity wave variances and squared amplitudes demonstrates that, in spite of the vertical averaging effect, gravity wave amplitudes determined in a 10 km vertical window are well suited for study of the interaction of the gravity wave distribution with the SAO winds.
Acknowledgements. We are grateful to ECMWF (http://www.ecmwf.int) for providing the ERA-Interim data. Further, we would like to thank NASA for making HIRDLS data available via their web page http://disc.sci.gsfc.nasa.gov/Aura/data-holdings/HIRDLS/index.shtml. Very helpful comments from the two anonymous reviewers are gratefully acknowledged.

The article processing charges for this open-access publication were covered by a Research Centre of the Helmholtz Association.

The topical editor C. Jacobi thanks two anonymous referees for help in evaluating this paper.

References


