Solar signal propagation: The role of gravity waves and stratospheric sudden warmings

I. Cnossen,¹ ² H. Lu,¹ C. J. Bell,³ L. J. Gray,³ and M. M. Joshi³

Received 26 May 2010; revised 15 October 2010; accepted 27 October 2010; published 27 January 2011.

[1] We use a troposphere-stratosphere model of intermediate complexity to study the atmospheric response to an idealized solar forcing in the subtropical upper stratosphere during Northern Hemisphere (NH) early winter. We investigate two conditions that could influence poleward and downward propagation of the response: (1) the representation of gravity wave effects and (2) the presence/absence of stratospheric sudden warmings (SSWs). We also investigate how the perturbation influences the timing and frequency of SSWs. Differences in the poleward and downward propagation of the response within the stratosphere are found depending on whether Rayleigh friction (RF) or a gravity wave scheme (GWS) is used to represent gravity wave effects. These differences are likely related to differences in planetary wave activity in the GWS and RF versions, as planetary wave redistribution plays an important role in the downward and poleward propagation of stratospheric signals. There is also remarkable sensitivity in the tropospheric response to the representation of the gravity wave effects. It is most realistic for GWS. Further, tropospheric responses are systematically different dependent on the absence/presence of SSWs. When only years with SSWs are examined, the tropospheric signal appears to have descended from the stratosphere, while the signal in the troposphere appears disconnected from the stratosphere when years with SSWs are excluded. Different troposphere-stratosphere coupling mechanisms therefore appear to be dominant for years with and without SSWs. The forcing does not affect the timing of SSWs, but does result in a higher occurrence frequency throughout NH winter. Quasi-Biennial Oscillation effects were not included.


1. Introduction

[2] Variations in solar ultraviolet (UV) irradiance that take place over the 11 year solar cycle are known to affect the upper stratosphere, where UV absorption by ozone takes place [Hood et al., 1993; Haigh, 1994, 1996; Gray et al., 2010]. Increased UV irradiance at solar maximum produces not only extra absorption directly, but also enhances the ozone concentration, making the equatorial upper stratosphere 1.5–2.5 K warmer compared to solar minimum [Hood, 2004; Crooks and Gray, 2005; Frame and Gray, 2010]. An even larger temperature signal associated with the 11 year solar cycle has been observed in the high-latitude regions of the lower stratosphere, and is particularly strong during winter [Lahitte and van Loon, 1988; Gray et al., 2004; Lu et al., 2009]. However, direct effects of changes in solar irradiance appear to be too small to cause these high-latitude signals [Hood, 2004; Gray et al., 2009]. It has therefore been proposed that solar UV forcing originating in the upper equatorial stratosphere may propagate dynamically poleward and downward during winter [Kodera and Kuroda, 2002; Matthes et al., 2004, 2006].

[3] Kodera and Kuroda [2002] proposed a propagation mechanism involving the redistribution of planetary wave forcing during Northern Hemisphere (NH) winter. They suggested that a region of anomalously strong westerlies in the subtropical upper stratosphere/lower mesosphere during solar maximum, in thermal wind balance with an enhanced pole-to-equator temperature gradient in the upper stratosphere due to enhanced equatorial heating, may deflect planetary waves poleward. The redistribution of planetary wave forcing towards higher latitudes further strengthens the polar vortex in the subtropics, so that the zonal wind there becomes even more westerly. As the area of anomalous westerly winds expands, this causes a further deflection of planetary waves, and so on, so that the westerly anomaly gradually moves poleward and downward.

[4] This mechanism was demonstrated in a simple, idealized modeling study by Gray et al. [2004], who imposed a
small easterly anomaly in the subtropical upper stratosphere in early winter, representative of solar minimum. This resulted in a consistently more disturbed winter with a weaker vortex and earlier sudden warming events compared to their unforced integrations. Matthes et al. [2004, 2006] successfully reproduced this same behavior in the NH winter using a more realistic full general circulation model, the Freie Universität Berlin Climate Middle Atmosphere Model (FUB-CMAM), and obtained a pattern of poleward and downward propagation of zonal wind anomalies, similar to observations. However, their modeled signal was much weaker than the observed signals. Also other studies with full chemistryclimate models and realistic solar forcing typically find signals that are weaker than seen in observations or at least at the lower end of observed ranges of peak responses [e.g. Marsh et al., 2007; Austin et al., 2008]. Matthes et al. [2004] argued that this could be due to the low variability produced by their model, as the amplitude of the response to solar forcing may be related to the amplitude of interannual variability [see also Kodera et al., 2003]. Kodera et al. [2003] further noted that the lower mesosphere zonal jet was not very well reproduced in these simulations, which they suggested may be due to the use of Rayleigh friction as a crude parameterization of gravity wave forcings in the FUB-CMAM.

[5] Rayleigh friction has long been the traditional approach to account for gravity wave effects, but is a rather crude method. It simply assumes that a drag must be present that is proportional to the ambient wind speed, with a height-dependent proportionality factor that is tuned such that a realistic climatology is obtained. It has several drawbacks. First, it is usually applied uniformly in time, latitude, and longitude, while real gravity wave sources and breaking events are likely to vary with location and be intermittent [Fritts and Alexander, 2003]. Second, it assumes that wave breaking always results in a drag on the mean wind, while in reality it may also accelerate winds in some cases, depending on the wave characteristics and background wind itself. Third, Rayleigh friction does also not conserve momentum [Shepherd et al., 1996; Shepherd and Shaw, 2004].

[6] A more sophisticated way to account for gravity wave effects that relieves some of these problems is by means of a gravity wave parameterization that incorporates some representation of the wave breaking process. This then determines where and when the waves break, and whether they strengthen or weaken the winds as they do so, depending on the wave characteristics and background winds. The first of such parameterizations were formulated by Lindzen [1981] and Dunkerton [1982], and since then have proven their value in many modeling studies [Fritts and Alexander, 2003]. It has been shown recently that the use of a gravity wave parameterization, as opposed to Rayleigh friction, can influence the modeled atmospheric response to a CO₂ forcing [Sigmond et al., 2008; Cnosson et al., 2009; Sigmond and Scinocca, 2010].

[7] There is some evidence that this may be the case for solar forcings as well. Shibata and Kodera [2005] compared the results obtained with a traditional Rayleigh friction approach in their model to those obtained with the parameterization described by Hines [1997]. They found that the Hines parameterization produced a more realistic semianual oscillation (SAO). Differences in the response to solar forcing between the Rayleigh friction and gravity wave scheme versions of their model were therefore interpreted as being due to the absence/presence of the SAO. Further, McCormack et al. [2007] showed that a reduction in Rayleigh friction strength resulted in a more robust solar cycle modulation of the quasi-biennial oscillation (QBO), due to a larger solar cycle variation in model dynamics near the stratosphere. These studies thus indicate that the representation of gravity wave effects can have consequences for the atmospheric response to solar forcing.

[8] The first objective of this study is to investigate such influences in more detail. We do this in an extension of the study of Gray et al. [2004], using a similar zonal wind forcing in the subtropical upper stratosphere. We use the Reading Intermediate General Circulation Model (IGCM3), which includes the troposphere, so that we can follow the response further down in the atmosphere and also assess tropospheric signals. In addition, our model integrations include a seasonal cycle, while Gray et al. [2004] modeled perpetual January conditions. The focus of our analysis will be on the poleward and downward propagation of the response in the NH, and we investigate how this propagation is affected by the representation of gravity wave effects in the model.

[9] Our study also builds on the work of Haigh et al. [2005] and Simpson et al. [2009] who investigated the tropospheric response to temperature forcings in the lower stratosphere. They used a simplified version of the model we use, with a lower top (at ~18.5 hPa compared to 0.1 hPa in our model) and no orography, and hence very weak planetary wave activity. They could therefore not study the downward propagation of the response to (an idealized) solar forcing from the upper stratosphere, where the solar UV and stratospheric ozone interaction takes place, to the lower stratosphere. We take their work a step further, by prescribing a forcing in the upper subtropical stratosphere.

This allows us to study the downward and poleward propagation of the responses within the stratosphere as well as the responses in the troposphere. We stress that we purely focus on dynamical propagation mechanisms, as do Haigh et al. [2005], Simpson et al. [2009], and Gray et al. [2004]. As a second, but related point of interest, we examine the role of stratospheric sudden warming (SSW) events in the response that is produced. SSWs occur in response to strong planetary wave forcings, which can cause a temporary breakdown of the winter polar vortex, with zonal winds reversing to easterlies and a corresponding warming as a result. As planetary waves are also thought to be involved in solar signal propagation, linkages between the occurrence of SSW events and the poleward and downward transport of solar signals may be expected.

Indeed, the modeling study by Gray et al. [2004] already showed that a subtropical forcing in the upper stratosphere affects both the timing and frequency of SSWs. Here we reexamine their results with a different model, including seasonality and the option to use a gravity wave scheme. We also look at the link between SSWs and solar forcing from another angle, by studying the effects of the absence/presence of SSWs on the responses obtained.

[10] Both the sensitivities to gravity wave forcing and SSW occurrence that we find in the responses are used as a
It accounts for a range of spectrum of gravity waves, and their 1981 and 1986, and a slab ocean and land Barnes 1990, 1991, providing a realistic simulation of the stratospheric mean state for physical processes via parameterization schemes, as described by Forster et al. 2000. These include a fast radiation scheme based on Morcrette 1990, 1991, a convection scheme based on Betts 1986, a boundary layer scheme based on Louis 1986, and a slab ocean and land surface scheme. In this sense, the IGCM3 is a GCM of intermediate complexity, bridging the gap between simple dynamical models and full state-of-the-art GCMs.

The model has been used extensively in radiative forcing studies [Forster et al., 2000; Joshi and Shine, 2003; Shine et al., 2003]. More recently, improvements to the model were made for the investigation of stratospheric processes and their effect on climate, in particular by providing a realistic simulation of the stratospheric mean state and variability [Bell et al., 2009]. The dynamical core of the model is the same as that of the model used by Haigh et al. [2005] and Simpson et al. [2009] for studying solar effects.

For this study, a T42 horizontal resolution (triangular truncation at wavenumber 42) was used with 38 vertical levels from 1000 to 0.1 hPa (16 levels in the troposphere, 19 levels in the stratosphere, and 3 levels above 1 hPa). Also, a gravity wave scheme was implemented and used for some of the simulations. The gravity wave scheme is based on work by Lindzen [1981] and Holton [1982], as described by Barnes [1990] and Joshi et al. [1995], and conserves momentum. It includes both orographic and nonorographic waves.

The orographic waves are assumed to be stationary (phase speed = 0 m/s) and have an amplitude based on the subgrid scale standard deviation of the topography, with a minimum value of 100 m, and the zonal wind speed at the surface. The mean orography from the U.S. Naval 1/6th degree resolution data set is used. A minimum value of 100 m was set to parameterize roughly residual breaking in the uppermost layers of the model of waves from slowly moving features (including over oceans). Two types of nonorographic waves, with phase speeds equal to the zonal wind speed at \( \sim 500 \text{ hPa} + 20 \text{ m/s} \) and \( \sim 500 \text{ hPa} – 20 \text{ m/s} \), respectively, and a fixed amplitude (a tunable parameter) are included. This is a simplified representation of nonorographic gravity waves compared to the much larger spectrum of waves that is often included in full climate models extending up to the mesosphere [e.g. Garcia et al., 2007]. These spectra sometimes have a latitudinal and seasonal dependence as well, while our two waves are distributed homogeneously in space and time, although some temporal and spatial dependence in phase speed is caused through variations in zonal wind speed at 500 hPa. Our simplified approach is justified for the following reasons. First, our model does not extend into the mesosphere, so that there is little gain in including waves with large phase speeds that would propagate through the stratosphere (and not break within the model domain). Second, Barnes [1990] notes that the overall impact of a relatively broad spectrum of waves on the zonal flow, even compared to just a single orographic mode, is not very large. And finally, the “true” spectrum of gravity waves, and their seasonal and latitudinal variation, remains not well known [e.g. Fritts and Alexander, 2003]. Including more waves, with or without a latitudinally and seasonally varying source, would thus not necessarily make our simulations more realistic.

2.2. Experimental Setup

Two control simulations were performed: one that used Rayleigh friction alone to account for gravity wave forcing and one that used the gravity wave parameterization; referred to hereafter as RF–C and GWS–C, respectively. When only Rayleigh friction was used, this was employed over the top six model levels (\( \sim 0.1–3 \text{ hPa} \)) with a time scale of 18 days at the lowest level, reducing to a time scale of 3 days at the top level, over which winds were relaxed toward zero. The overall strength of the Rayleigh friction was chosen such that the zonal mean temperature and zonal wind climatologies for RF–C matched as closely as possible observed climatologies and climatologies obtained with GWS–C. This excludes as much as possible the effects of different background climatologies on the response to the forcing. When the gravity wave scheme was used, a weak Rayleigh friction was still retained at the top four model levels (\( \sim 0.1–1 \text{ hPa} \)) with a time scale of 19 days at the lowest level reducing to 4.7 days at the top level, to avoid spurious wave reflections from the model top. The Rayleigh friction timescales for the RF–C and GWS–C simulations are shown in Figure 1.

Two perturbed simulations were also performed and will be referred to as RF–P and GWS–P. The perturbation
Figure 3. Distribution of the occurrence frequency of stratospheric sudden warming (SSW) events for the RF-C and GWS-C simulations and the observed SSW frequency from Charlton and Polvani [2007] (CP07) based on the average of the National Centers for Environmental Prediction (NCEP) and ERA-40 occurrences.

3. Results

3.1. Control Simulations

Figure 2 shows the October-December (OND) zonal mean temperature and zonal wind climatologies, together with their standard deviation (shaded), for the GWS-C (left) and RF-C (middle) simulations. The difference between the climatologies is also shown (right), with the shading now indicating statistical significance. The climatologies for the two simulations are very similar (which they were designed to be) and capture the main features of the observed zonal mean temperature and wind structures. There are statistically significant differences between the simulations, mainly in the stratosphere, but these differences are relatively small and, in particular, the zonal wind in the NH polar vortex matches to a high degree.

Figure 2. (top) Zonal mean temperature and (bottom) zonal wind climatologies for the (left) GWS-C simulation, (middle) the RF-C simulation, and (right) the difference between the GWS-C and RF-C simulations for October-December (OND). For Figure 2 (left and middle) the shading indicates the standard deviation. For Figure 2 (right) the shading indicates statistical significance at the 95% (light shading) and 99% (dark shading) level, as determined with a t test.
3.2. Perturbed-Control Differences

[21] Figure 4 shows sequences of differences between the perturbed and control integrations for gravity wave simulations in zonal mean temperature (left) and zonal wind (right) for 10 day intervals from the start of November to the start of December. The temperature signal clearly moves poleward and downward from 1–10 to 21–30 November, until it nearly vanishes in 1–10 December. The zonal wind signal propagates in conjunction with the temperature signal, and there is also a response present in the troposphere throughout November, consisting of a strengthening of zonal winds at 40°N and a weakening at 60°N. This roughly maps on to the Northern Annular Mode (NAM) [Thompson and Wallace, 1998], although we note that the modeled zonal wind response to our forcing is shifted ~5° northward compared to the NAM dipole pattern in observational data, and that the modeled tropospheric jet is located up to 5° further north than the observed jet.

[22] The poleward and downward movement of the temperature and zonal wind responses in the stratosphere is dynamically consistent with the differences in EP flux and EP flux divergence, as shown in the left-hand panels of Figure 5. The right-hand panels of Figure 5 show the transformed Eulerian mean (TEM) [Andrews and McIntyre, 1976; Andrews et al., 1987] residual circulation, which is a measure of the large-scale meridional circulation that is induced by wave forcing, that is, the Brewer-Dobson (B-D) circulation in the stratosphere.

[23] In 1–10 November there is enhanced EP flux into the equatorward flank of the polar vortex, and enhanced EP flux convergence (near 40°N, 1 hPa). This signifies an increase of the wave forcing on the mean flow, which acts to reduce the strength of the polar vortex. The enhanced wave forcing strengthens the B-D circulation in the area below, with a corresponding weakening in the region above the enhanced wave forcing, effectively moving the circulation downwards. During 11–20 and 21–30 November, the upward EP flux into the polar vortex is further enhanced, while the centre of the activity moves poleward. As a result the polar vortex is gradually being weakened from the outside equatorward flank to its inner core. The pair of positive and negative anomalies in the B-D circulation also moves poleward (from ~15°N in 1–10 November to ~60°N in 21–30 November), although the negative anomaly is no longer significant by 21–30 November. In 1–10 December the system is recovering, but slightly overshooting, as the EP flux is now reduced and EP flux divergence is enhanced compared to the control run. The relatively weaker planetary wave forcing allows the polar vortex to strengthen again and weakens the high latitude branch of the B-D circulation. After 1–10 December (not shown), there is very little significant signal remaining, indicating that the system has largely returned to its equilibrium state. This is in agreement with the observations of Lu et al. [2009] that the NH signals in the polar stratosphere have a life span of ~30–50 days, which is approximately the thermal relaxation time scale in the lower stratosphere [Newman and Rosenfield, 1997].

[24] The responses obtained from the RF simulations are of similar strength and show similar spatial patterns in the stratosphere, but there are some differences in the timing of their poleward and downward propagation. The RF response moves noticeably quicker towards the pole, reaching the pole ~10 days earlier than the GWS response, with most of the downward propagation occurring after that time.

[25] In the troposphere, significant responses in the zonal wind are detected, which are substantially different for RF and GWS, as shown in Figure 6. There is relatively little movement of the tropospheric signal over time, and it is most significant for November. Therefore, only an average response for November is shown. The zonal wind response for GWS consists of a weakening of the zonal flow at 10–20°N at 100–200 hPa, a strengthening throughout the depth of the troposphere centered at 40°N, and a weakening around 60°N at 250–500 hPa. This means that the core of the jet is strengthened (see Figure 2 for jet position in the control simulations). The RF zonal wind signal consists instead of a weakening from 15–20°N to 30–35°N through the depth of the troposphere and a slight strengthening around 50°N. This represents a poleward movement of the tropospheric jet. In both cases, the tropospheric responses do not appear to have directly descended from the stratosphere. They remain roughly in place throughout November-December, regardless of the temporal evolution of the stratospheric responses. At certain times they are therefore opposite in sign to the stratospheric responses.

3.3. Effects on Timing and Frequency Distribution of SSWs

[26] Because the GWS-C simulation produced an SSW distribution that is in better agreement with observations than the RF-C simulation produced, the following sections will make use of the GWS results only. Figure 7 compares the SSW distributions for the forced and control simulations. The overall shape of the SSW distribution remains the same, but the forced run shows a consistent increase in SSWs throughout winter compared to the control runs. This is in agreement with the finding by Gray et al. [2004] that the rest of the NH winter becomes more disturbed as a result of an easterly forcing in early winter in the subtropical upper stratosphere. However, a change in the timing of SSWs, as found by Gray et al. [2004], cannot clearly be seen in our results.

3.4. SSW Effects on the Responses

[27] To investigate the influence of SSWs on the tropospheric responses to the forcing, we separated the data for the forced and control GWS simulations in years with and without SSWs during OND. For both simulations, 12 years with SSWs and 38 years without SSWs were identified (this included three years with an October SSW for GWS-C, which were not shown in Figure 7). As noted previously, the zonal wind response in the troposphere resembles a NAM-like pattern. Therefore a NAM-like index was calculated to show the difference in the temporal evolution of the signal between the non-SSW and SSW data sets. This index was calculated by subtracting the area-weighted average of the geopotential height for 40–60°N from the area-weighted average for 60–90°N. These latitude bands were chosen specifically to capture the responses most clearly. The geopotential height difference was then normalized by subtracting the mean climatology for the control and per-
Figure 4. Differences in the (left) zonal mean temperature and (right) zonal wind climatologies between the GWS perturbed (GWS-P) and GWS control (GWS-C) simulations for 10 day averages from 1–10 November to 1–10 December. Light shading indicates 95% statistical significance and dark shading indicates 99% statistical significance, as determined with a $t$ test.
turbed data sets, and dividing the result by the standard deviation of the combined data sets. Positive values correspond to a positive NAM phase and negative values to a negative NAM, although our index is not directly comparable to the NAM due to the different latitude bands used. Figure 8 shows pressure-time sections of the response in the normalized geopotential height difference between 60°–90°N and 40°–60°N for the data sets with and without SSWs.

For years with SSWs, stronger poleward propagation of the response to the forcing takes place in the stratosphere, so that the positive response in the stratospheric geopotential height difference occurs earlier and is stronger as well. However, the overall propagation pattern of a positive signal following a negative signal is similar for years with and without SSWs in the stratosphere. In contrast, in the troposphere we find a negative signal around 10 November that appears to have descended from the stratosphere for years with SSWs, while we find a signal of the opposite sign around the same time, and in fact throughout the whole time interval presented, when years with SSWs are excluded. The non-SSW tropospheric response does not appear directly connected to the stratospheric response. When the entire data set (SSW + non-SSW years) is processed, the signal around 10 November is weaker, while the signal around 20 November is stronger. The signals from the two conditions thus act to cancel each other out partly around 10 November, while they add up around 20 November.

4. Discussion

4.1. Influences of Gravity Wave Effects on Solar Signal Propagation

Previous modeling studies have demonstrated that the representation of the stratosphere in a GCM can influence the troposphere [e.g., Boville, 1984; Song and Robinson, 2004; Sigmond et al., 2008]. Here we build on those findings and show a specific influence of the representation of gravity wave effects on the downward and poleward propagation of an idealized solar forcing in the upper stratosphere. Differences between the GWS and RF simulations occur in terms of the timing and extent of poleward and downward propagation of the responses in the stratosphere, and also, in general, in the responses in the troposphere, even though the differences between the RF and GWS simulations in terms of forcing and gravity wave representation are in the (upper) stratosphere. The influences we find in the lower stratosphere and troposphere are thus indirect.

The tropospheric zonal wind response for GWS is to some extent in agreement with observations by Haigh et al. [2005], who found that the tropospheric jets are weaker and
shift poleward for higher solar activity. The GWS results did not show the change in jet position, but did show a change in strength in the same sense, bearing in mind that our results are presented as solar minimum-maximum conditions.

The RF results on the other hand showed mostly a change in jet position, but it moved poleward for solar minimum conditions compared to solar maximum, that is, in the opposite sense to the observations by Haigh et al. [2005]. For both RF and GWS the strength of the tropospheric zonal wind signal is comparable to the strength of the observed anomalies reported by Haigh et al. [2005]. The GWS pattern of a strengthening of the jet at 40°N with a weakening on either side also matches the general pattern of the zonal wind signal observed by Frame and Gray [2010], although their pattern is centered at 30–35°N, while the RF pattern does not match. The GWS simulations thus give a more realistic tropospheric response, in better agreement with observations.

[31] Sigmond and Scinocca [2010] found that the sensitivity of the doubled CO$_2$ response to parameterized orographic gravity wave drag [Sigmond et al., 2008] was largely due to differences in the control climatologies. In our case, the differences in response to an idealized solar forcing do not seem to be related to such differences, as the control

**Figure 7.** Distribution of the occurrence frequency of SSW events for the GWS for the control and perturbed simulations.

**Figure 8.** Pressure-time section of the normalized geopotential difference between 60°N–90°N and 40°N–60°N between the GWS perturbed and GWS control data sets (top) without SSWs and (bottom) with SSWs from 10 October to 30 December. There are 12 years with SSWs and 38 years without SSWs in both GWS-C and GWS-P. Light shading indicates 90% statistical significance and dark shading 95% statistical significance, as determined with a $t$ test.
climatologies were designed to be similar. We can also exclude the possibility that the differences are related to different distributions of SSW events (which were shown to differ more substantially), as excluding years with SSW events in OND did not result in better agreement between the RF and GWS responses in OND (results not shown). Shibata and Kodera [2005] interpreted the differences in response they found between their model versions with Rayleigh friction and the Hines parameterization scheme in terms of the absence/presence of an SAO. In our simulations however, we do not get a realistic SAO in either the RF or GWS simulations. Therefore, while the SAO may have an influence in reality, it is not the reason for the differences we find in our results. We will now explore two alternative explanations.

32 First, it may be possible that the gravity wave forcing itself played a role in the propagation and maintenance of the signal, by redistributing gravity wave momentum deposition, similar to how planetary waves are thought to play a role in signal propagation by redistributing their momentum deposition. We therefore examined the difference in gravity wave-induced accelerations between the forced and control GWS runs, which revealed that these always acted to reduce the zonal wind responses found (results not shown). So rather than amplifying the response, the direct effects of changes in gravity wave momentum deposition acted to diminish the response, and we can eliminate this possibility.

33 A second pathway for gravity wave effects to influence the response to our forcing is through indirect effects on planetary waves. As noted, the EP fluxes and EP flux divergence were about 25% larger in the NH upper stratosphere during OND for GWS-C compared to RF-C. This is consistent with the finding by McLandress and McFarlane [1993] that longitudinal variations in gravity wave drag (which would be missing in a Rayleigh friction approach) can enhance planetary wave amplitudes and EP flux divergence. Considering the mechanism proposed by Kodera and Kuroda [2002], this could explain why the stratospheric response propagate differently for the GWS and RF versions of the model, although the detail of the differences is not straightforward to explain.

34 The stronger planetary wave activity may also be responsible for the slightly stronger interannual variability found for GWS. However, this small enhancement in interannual variability is not sufficient to determine whether a lack of interannual variability is responsible for a too weak response to solar forcing, as suggested by Matthes et al. [2004] and Kodera et al. [2003]. In terms of strength, the responses obtained with the RF and GWS versions of the model are very similar.

35 In conclusion, our results indicate that it is important to model the planetary wave activity correctly, and as gravity wave effects can modify this activity substantially, a more realistic representation of gravity wave effects seems to be necessary to achieve this. The use of a gravity wave scheme does not only affect stratospheric responses, but also the responses in the troposphere, which become more realistic when the gravity wave scheme is used. We note that our model simulations, despite making use of a gravity wave scheme, still do not necessarily provide a realistic description of gravity wave effects, as strong assumptions were made on the characteristics of the waves. However, more information on the global distribution of gravity wave effects is becoming available [e.g. Alexander et al., 2008] and future studies that attempt to realistically model the response to solar forcing should take advantage of this.

4.2. Influences of SSW Events on Solar Signals

36 Separating our data into years with and without SSW events in OND showed that responses are substantially different under both conditions. Years with SSWs are highly disturbed and typically have a strong planetary wave forcing in the high latitude upper stratosphere, causing the polar vortex to break down. Again, considering the Kodera and Kuroda [2002] mechanism of solar signal propagation, we would expect poleward and downward propagation to be enhanced under such conditions and that is indeed what we observe. During years with SSWs a negative stratospheric signal appears to descend down into the troposphere directly, becoming strongest there around 10 November. This is followed by a positive stratospheric response, which also appears to descend down into the troposphere, although the tropospheric response, while positive also, is not significant. In contrast, when years with SSWs are excluded, the tropospheric signals are positive throughout and do not change in conjunction with the time-varying changes in the stratosphere. The tropospheric signals appear therefore more disconnected from the stratospheric signals in this case.

37 This does not mean that there is no connection between the phase of the NAM in the troposphere and stratosphere when no SSWs are present, but suggests that different types of troposphere-stratosphere coupling are dominant for disturbed (with SSWs) and quiet (no SSWs) conditions. The mechanism responsible for the non-SSW signals is unlikely to be a direct tropospheric extension of the mechanism responsible for the poleward and downward propagation of the stratospheric signals. In other words, those tropospheric signals do not appear to be due to a change in planetary wave forcing, and an associated change in a mean meridional circulation, extending from the stratosphere into the troposphere, as this should have resulted in signals of the same sign in both the stratosphere and troposphere. We therefore discuss two alternative coupling mechanisms.

38 The first mechanism involves changes in the tropospheric mean meridional circulation, forced by changes in eddy momentum fluxes associated with synoptic waves within the troposphere [Simpson et al., 2009]. The changes in eddy momentum fluxes are brought about by changes in temperature gradients and zonal wind accelerations at the tropopause, in response to changes in the lower stratospheric temperature structure and flow. Simpson et al. [2009] demonstrated the above mechanism by modeling the effect of an altered latitudinal temperature structure in the lower stratosphere on the troposphere. They showed that the response slowly develops in the upper troposphere, over about 10 days, and gradually spreads to the lower troposphere, with the main response structure established after about 20 days.

39 In our results we do not see a clear downward movement of the signal from the upper troposphere into the lower troposphere over time, even though our model is a more advanced version of the model used by Simpson et al.
[2009]. Also, in our experiments the stratospheric signal does not penetrate down into the lower stratosphere when years with SSWs are excluded, while there is still a clear tropospheric response. It is therefore unlikely that the above mechanism is responsible for our tropospheric response, although it is still possible that tropospheric eddies are involved in amplifying/maintaining the tropospheric response [see also Kushner and Polvani, 2004; Song and Robinson, 2004], once initiated by some other process.

[40] The second mechanism involves changes in the reflection of planetary waves by the stratosphere back into the troposphere, where they are subsequently absorbed [Perlwitz and Harnik, 2003, 2004; T. A. Shaw et al., Downward wave coupling between the stratosphere and troposphere: The importance of meridional wave guiding and comparison with zonal-mean coupling, submitted to Journal of Climate, 2010]. Perlwitz and Harnik [2004] argued that this mechanism becomes more important when the polar vortex is strong, in particular the lower part of the vortex (∼30 hPa), as more wave activity in that case is reflected back into the troposphere, rather than being absorbed by the stratosphere. They also found that it is important mainly on short time scales (up to 12 days).

[41] This mechanism could potentially explain why we find a different tropospheric signal for years with and without SSWs. During years with SSWs, troposphere-stratosphere coupling via reflection of planetary waves would be weak, while the mechanism would be more important, resulting in stronger coupling, during years without SSWs, when the polar vortex is stronger. This is consistent with our finding that tropospheric signals of opposite sign appear or become enhanced for years without SSWs, while no significant signal, or a signal with the same sign as that in the stratosphere appears for years with SSWs.

[42] Previous studies have shown that there is a strong coupling between the troposphere and stratosphere associated with SSWs [Baldwin and Dunkerton, 2001; Charlton and Polvani, 2007], but this appears to be a more direct coupling, with signals from the stratosphere apparently descending down into the troposphere. This is indeed what we observe in our results with SSWs, although most of the responses are not significant. This could be due to the fact that the data sets with SSWs are much shorter than the data sets without SSWs, and within the SSW data sets the SSWs also occurred at different times. This gives a noisier data set and makes it harder to establish a significant response. A second possibility is that our forcing during years with SSWs could be relatively less important to the troposphere than it is during years without SSWs, due to the dominance of the SSW influence on the troposphere-stratosphere coupling in years with SSWs.

[43] Finally, we note that the tropospheric responses of the SSW and non-SSW are sometimes of the opposite sign and can therefore act to cancel each other out. This could be a possible reason for difficulties in establishing a significant tropospheric signal in observational data, as both disturbed and undisturbed years are normally included.

4.3. QBO Influences

[44] The interaction between the QBO and solar forcing is still unclear. Some studies have found that solar forcing influences the QBO [McCormack et al., 2007] or that the QBO affects solar signals [Labitzke and van Loon, 1988; Gray et al., 2004; Labitzke et al., 2006], while others find little interaction [Austin et al., 2008]. Lu et al. [2009] and Ito et al. [2009] found that the effects of the QBO on solar signal propagation occur predominantly in late winter.

[45] Our results were obtained with permanently weak easterlies in the equatorial stratosphere, so that any interactions between solar forcing and the QBO have not been considered. However, as we focus on early winter, the absence of a QBO may not have had a large effect. Still, the QBO affects the distribution of planetary wave activity, which plays a key role in the poleward and downward propagation of the signal in the stratosphere. Inclusion of a realistic QBO could therefore in principle modify our results somewhat. On a background of westerly winds in the equatorial stratosphere, an easterly forcing should still have the effect of deflecting waves poleward, but the strength of the signal and the timing and extent of poleward propagation might be different.

5. Summary and Conclusions

[46] Our results broadly confirm the mechanism for solar signal propagation in the stratosphere proposed by Kodera and Kuroda [2002]. We find that, in agreement with their theory, the redistribution of planetary wave activity can strengthen an initial forced signal, and transport it poleward and downward from the equatorial upper stratosphere. We find that the type of representation of gravity wave effects in our model influences this process, changing the timing and extent of poleward and downward signal propagation in the stratosphere. This takes place most likely through indirect effects of gravity wave-induced accelerations on planetary waves. The results obtained with the gravity wave scheme are more realistic than those obtained with Rayleigh friction, as they are in better agreement with observed solar signals, in particular in the troposphere.

[47] The GWS results also produce a more realistic distribution of SSW events. The absence/presence of SSW events has an effect on the propagation of the response to our forcing, mainly in the troposphere. For years with SSWs tropospheric signals appear to descend directly from the stratosphere, while they appear more disconnected when SSW years are excluded. We suggest that this is due to different types of troposphere-stratosphere coupling being active under conditions with and without SSWs. Under quiet conditions, a signal in the troposphere of the opposite sign to that in the stratosphere may be generated through small modifications in the reflection of planetary waves back into the troposphere. Once initiated, this signal may be maintained and/or strengthened locally through changes in eddy momentum fluxes. In contrast, under disturbed conditions, when SSWs occur, the troposphere-stratosphere coupling occurs more directly, and the tropospheric response is an extension of that in the stratosphere.

[48] The forcing also increases the number of SSWs, but does not influence their timing, as found by Gray et al. [2004]. We therefore confirm only part of the findings of Gray et al. [2004]. Their result that earlier SSWs occurred when an easterly forcing in the subtropical upper strato-
sphere was applied may have been related to the absence of a seasonal cycle in their model integrations.

[49] Acknowledgments. We are very grateful to three anonymous reviewers for their constructive comments, which helped to improve the original manuscript.

References


J. Climate, 17, 629–639.


C. J. Bell, L. J. Gray, and M. M. Joshi, Department of Meteorology, University of Reading, Reading RG6 6BB, UK.

I. Cnossen, High Altitude Observatory, National Center for Atmospheric Research, 3080 Center Green Dr., Boulder, CO 80301, USA. (icnossen@ucar.edu)

H. Lu, British Antarctic Survey, High Cross, Madingley Rd., Cambridge CB3 0ET, UK.