Characterization of the 11-Year Solar Signal Using a Multiple Regression Analysis of the ERA-40 Dataset

SIMON A. CROOKS
Atmospheric, Oceanic and Planetary Physics, University of Oxford, Oxford, United Kingdom

LESLEY J. GRAY
Department of Meteorology, Reading University, Reading, United Kingdom

(Manuscript received 19 April 2004, in final form 3 September 2004)

ABSTRACT

A multiple linear regression analysis of the ERA-40 dataset for the period 1979–2001 has been used to study the influence of the 11-yr solar cycle on atmospheric temperature and zonal winds. Volcanic, North Atlantic Oscillation (NAO), ENSO, and quasi-biennial oscillation (QBO) signatures are also presented. The solar signal is shown to be readily distinguishable from the volcanic signal. The main solar signal is a statistically significant positive response (i.e., warmer in solar maximum) of 1.75 K over the equator with peak values at 43 km and a reversed signal of similar magnitude at high latitudes that is seasonally dependent. Consistent with this is a statistically significant zonal wind response of up to 6 m s$^{-1}$ in the subtropical upper stratosphere/lower mesosphere that is also seasonally dependent. The wind anomalies are westerly/easterly in solar maximum/minimum. In addition, there is a statistically significant temperature response in the subtropical lower stratosphere that shows similarity in spatial structure to the QBO response, suggesting a possible interaction between the solar and QBO signals in this region. The solar response in tropospheric zonal winds is small but significant, confirming previous studies that indicate a possible modulation of the Hadley circulation.

1. Introduction

It has been suggested that changes in solar irradiance over the 11-yr solar sunspot cycle influences the Earth’s atmosphere. Characterization of the impacts of this influence and possible mechanisms is increasing but it is, generally, still extremely poor in comparison to that of other climate forcing agents (Houghton et al. 2001) despite solar influences being the primary source of energy in the global climate system. One of the fundamental problems in attributing atmospheric changes to changes in solar activity is that the total solar irradiance (TSI) variations from the minimum to the maximum of a solar sunspot cycle are only of approximately 0.1%, and it is controversial how such a small change in the already noisy solar irradiance measurements manifests itself as changes in the earth’s atmosphere. Earlier studies of this issue (e.g., Friis-Christensen and Lassen 1991) were criticized for being based upon correlations of heavily filtered time series and the failure to provide a plausible physical mechanism to amplify the response to these relatively small changes in TSI.

A number of regression studies have been performed using various temperature datasets to explore the 11-yr solar cycle. Pioneering work was performed by La-bitzke and van Loon (1988), who discovered significant correlations between high latitude stratospheric temperatures/geopotential heights and solar ultraviolet (UV) when the data was stratified according to the phase of the quasi-biennial oscillation (QBO) in lower-stratospheric equatorial winds, but not when the full dataset was used in their analysis. Since then several other observational studies have been performed. In a regression analysis using 18 years (1980–97) of Stratospheric and Microwave Sounding Units(SSU/MSU) temperature data corrected and compiled by the National Centers for Environmental Prediction (NCEP)–Climate Prediction Center (CPC), Hood (2004, referred to hereafter as H2004) found a vertical three-cell pattern of temperature response in the stratosphere at low latitudes. Assuming an equivalent of 130 units change in 10.7-cm flux over an average solar cycle, he found a positive temperature response maximizing at about 48 km (the highest level in their dataset) at over 2 K per solar cycle. This decreased through zero (34 km) to $-1$ K at 32 km and was partially statistically significant. Below this (16–20 km) a second significant positive response (1–1.5 K) was present, thus forming
the three-cell pattern. However, establishing temperature trends and signals in this dataset, especially above 10 hPa, is difficult due to spurious discontinuities introduced by changes in satellite instruments (see Ramaswamy et al. 2001).

Scaife et al. (2000, hereafter S2000) used SSU/MSU data, which was adjusted and compiled in a slightly different way, and also regressed the temperature estimates for 1979–97 onto the 10.7-cm solar radio flux and a linear trend to obtain a pattern of solar induced response (see also Keckhut et al. 2005). Although they also found a three-cell pattern of temperature response in equatorial regions, there were significant differences from the signal found by H2004. S2000 found a negative response above the stratopause with a zero response at about 49 km, corresponding to the region of large positive response found in the CPC data of H2004. Their upper-level positive maximum was lower than H2004 at about 40 km instead of 48 km, with a much smaller amplitude between 0.5 and 0.75 K per 100 units of 10.7-cm radio flux. Below this, a local negative minimum was found centered at about 22 km but with no statistical significance. Thus, the greatest difference between the two results is the height and magnitude of the upper-level maximum and the discrepancy in sign in the mid stratosphere (25–35 km). A further difference between the two analyses is the latitudinal structure of the response. While the gross structure was similar, with the maximum response between ±30° latitude (irrespective of sign), the S2000 response was negative in the lower stratosphere at latitudes greater than about 40° in both hemispheres. The final result is two very different patterns of possible solar-induced temperature response. These inconsistencies, and subsequent uncertainties, provide motivation for the analysis presented in this paper.

Importantly, previous studies have neglected to compensate for volcanic eruptions in their analyses. Salby and Shea (1991) pointed out that this could be problematic since atmospheric behavior attributed to solar variations could actually be due to volcanic eruptions. Hence, in this study great care is taken to avoid this pitfall by attempting to distinguish between these two (and several other) sources of atmospheric variability.

Haigh (2003) also performed a linear multiple regression of zonal mean temperatures from the NCEP–National Center for Atmospheric Research (NCAR) reanalyses but concentrated primarily on the analysis of the tropospheric response. In addition to the positive response in the lower equatorial stratosphere, there were also statistically significant vertical bands of positive response in the troposphere in both hemispheres between 20° and 60°.

In addition to these global analyses of the 11-yr solar signal, various localized datasets have been analyzed, including rocketsonde, radiosonde, and lidar data (Dunkerton et al. 1998; Angell 1991; Keckhut et al. 2005). While all of these analyses have confirmed an 11-yr solar signal in the data, there is nevertheless no consensus yet on the exact height and latitude structure of the temperature response.

In this paper we carry out a multiple regression analysis of the new European Centre for Medium-Range Weather Forecasts (ECMWF) 40-Yr Reanalysis (ERA-40) dataset (see section 2). Various components of stratospheric temperature are disentangled to illustrate that each individual forcing scenario produces an individual pattern of response relatively distinct from the others. We employ the zonally averaged monthly mean data covering the period 1979–2001, thus encompassing two 11-yr solar cycles. Although data from the full length (mid-1957–2001) of the dataset were available, we choose to use the period since 1979, when reliable satellite data for the middle atmosphere became available. Before this, the upper-stratospheric and mesospheric fields of the analyses are inevitably dominated by the model, which does not contain any mechanism to represent the 11-yr solar cycle. However, the inclusion of the earlier data produces similar, yet weaker, and less significant results. The same conclusion was found by Haigh (2003) and Labitzke et al. (2002a).

The solar proxy used in this, as in most studies, is the 10.7-cm radio flux, which spans the longest time period. More physical and realistic proxies do exist and would include the He I line, the Mg II line, total solar irradiance, and UV irradiance. The use of time series representative of UV radiation is an obvious choice due to the important role that it plays in driving middle atmosphere ozone chemistry, but as yet no single instrument has measured the solar UV irradiance over even one complete cycle. The joining together of data from various satellite instruments would inevitably lead to discontinuities and these must be avoided wherever possible. On the basis of previous studies that have tested different proxies (e.g., Donnelly et al. 1986; Keckhut et al. 1995), the 10.7-cm solar flux, which closely tracks the temporal behavior of the UV changes on 11-yr time scales (Keckhut et al. 2005), is taken in our analysis to represent solar variability. S2000 and others have used the smoothed 10.7-cm flux. However we find that smoothing the 10.7-cm flux time series with a 29-month running mean makes very little difference to our results, implying that the solar response we see is not necessarily dependent upon high frequency variability.

A strength of examining a sophisticated analyses, such as ERA-40, is that it incorporates data from various observation sources and allows a complete global coverage. Herein lies a possible problem since, if the observational data is sparse over certain regions or in time, then there is an increased chance that the details of the numerical model and/or the assimilating method could influence the results. As previously discussed, attempts have been made to avoid this problem, especially in the stratosphere, by only analyzing over time periods that are rich in satellite data.
The analysis presented here is able to reinforce and further the study of Haigh (2003), whose dataset has an upper limit at 10 hPa, by also examining the solar response in the upper stratosphere and the lower mesosphere in addition to the lower atmosphere. A spatially complete zonal mean temperature and zonal wind response to the 11-yr solar cycle from the surface to 0.1 hPa is produced. Spatial patterns with this vertical (and horizontal) coverage derived using data that encompasses the last two 11-yr solar cycles are not available from any other climatological dataset.

Section 2 summarizes our data and analysis techniques. Section 3 discusses both the stratospheric and tropospheric annual and seasonally averaged solar signal in the ERA-40 zonal mean temperature and zonal wind fields. For completeness, section 3 also documents the zonal mean temperature and zonal wind signals of volcanic activity, ENSO, the North Atlantic Oscillation (NAO), and the OBO. Finally the results from the analysis are summarized in section 4.

2. Data and methods

The ERA-40 data assimilation system uses the Integrated Forecasting System (IFS) developed jointly by ECMWF and Météo-France. A three-dimensional variational method is used to assimilate the observations into the spectral model, which has 60 vertical levels and T159 horizontal spectral resolution. Data both on the model levels and interpolated onto standard pressure surfaces are available. We use monthly averaged data derived from the 6-hourly analyses on standard pressure levels below 100 hPa (viz., 100, 150, 200, 250, 300, 400, 500, 600, 700, 775, 850, 925, and 1000 hPa). Above this, model level data have been used for improved accuracy in the middle and upper stratosphere. These correspond approximately to the following pressure surfaces: 0.1, 0.3, 0.5, 0.8, 1.2, 1.6, 2.1, 2.7, 3.4, 4.2, 5.2, 6.4, 8.0, 9.8, 12, 15, 19, 23, 29, 36, 44, 55, 67, and 80 hPa.

Prior to 1972 the atmospheric measurements available to ERA-40 are limited mainly to radiosondes. However, since then the reanalysis makes comprehensive use of satellite data, starting from the early Vertical Temperature Profile Radiometer (VTPR) in 1972–79, then later (from 1979) the Television Infrared Observation Satellite (TIROS) Operational Vertical Sounder (TOVS) [consisting of SSU/High Resolution Infrared Radiation Sounder (HIRS)/MSU, the Special Sensor Microwave Imager (SSM/I; from 1987), the European Remote Sensing Satellite (ERS; from 1991)], and the Advanced TIROS Operational Vertical Sounder (ATOVS) data [from 1998, consisting of HIRS and the Advanced Microwave Sounding Unit (AMSU)]. Cloud Motion Winds were also used from 1979 to the present time. The ozone observations used in the ERA-40 are Total Ozone Mapping Spectrometer (TOMS) total ozone, Solar Backscatter Ultraviolet Instrument (SBUV) ozone layer measurements, and TOVS/HIRS channel-9 radiances; all are available from 1978 onward. Note, in contrast to ERA-15 where retrievals of temperature and humidity were assimilated, ERA-40 assimilates radiances directly, thus avoiding a bias toward climatologies. (A more detailed examination of the ERA-40 dataset can be obtained online at http://www.ecmwf.int/research/era/.)

A standard linear multiple regression technique using an autoregressive model (AR) of order 3 was used to separate various climatic effects in the temperature and zonal wind datasets. The data were deseasonalized by removing the climatology derived from 1979 to 2001, before fitting them to 20 independent indices via an ordinary, linear least squares regression model. Twelve of the indices correspond to a series of square pulses superimposed upon a zero line (square waves) with each pulse representing a single month and each individual index containing pulses pertaining only to that particular month. Thus, these indices, by definition, will be orthogonal and together should represent any recurrent annual variations in the other indices, and are necessary in the formulation of statistical tests. An alternative method of deseasonalizing the variables in the regression would be to represent any annual and semi-annual cycles by sine and cosine waves (four in total). However, the former method was chosen to avoid the risk of aliasing between the sine and cosine terms and any sinusoidal behavior in the other indices, and also to ensure the correct number of degrees of freedom in statistical tests. However, when no seasonal cycle is present in the dependent variable, such as in the analysis of individual seasons or months, these 12 orthogonal terms are no longer necessary and are hence excluded.

The regression model also incorporated a linear trend term, the solar 10.7-cm radio flux, the stratospheric aerosol optical depth time series (related to volcanic eruptions) from Sato et al. (1993; updated data are available online at http://www.giss.nasa.gov/data/strataer/), the NAO index from Jones et al. (1997), a Cold Tongue Index (highly correlated to ENSO) from the University of Washington, Seattle (see online at http://tao.atmos.washington.edu/data_sets/cti), and two QBO time series as discussed below. The final index was a step function (at the end of March 1986) to account for a discontinuity in the ERA-40 reanalyses due to the joining together of two independently run analysis streams (A. Untch 2003, personal communication). Inevitably the reanalyses could contain other spurious discontinuities due to changes in satellites; however, the exact timing of these changes is not obvious, hence no further step functions are introduced. Moreover, the step before April 1986 is largely redundant since it had no effect on the response patterns except the temperature trend, which is not presented here.

The use of an autoregressive noise model in the analysis ensures that we do not overestimate the confidence in our response amplitudes. The order 3 was a
semi-arbitrary decision since the response patterns are
more or less indifferent to the order \( p \) of the AR model
(provided that \( p > 0 \)). However, since the data contains
a QBO component (in the tropical stratosphere), if not
all of it is fitted in the regression model, the resulting
residual will exhibit periodic behavior. It is therefore
appropriate to fit the residual using an AR model with
more than one degree of freedom (that is, \( p > 1 \)) to
take account of periodicity in the residual greater than
one time step. Hence a conservative estimate of three
was used, but the results were not sensitive to this
choice of value.

Figure 1 shows the time evolution of the standardized
physical indices fitted in the regression model. The two
orthogonal QBO time series employed were the first
two principal components (PCs) of the residual of a
regression of the ERA-40 zonally averaged strato-
spheric zonal wind onto the other 18 indices: 10.7-cm
solar flux, optical depth, NAO index, ENSO, trend, the
step function, and the 12 square pulse time series
(square waves) to compensate for any annual cyclic be-
havior in the indices and the data (although this is not
necessary for the latter since it has deseasonalized prior
to the regression). By simultaneously fitting a red-noise
model of order 3 in the regression the QBO time evo-
lution is represented strikingly well by the residual. An
alternative approach would be to use the QBO equa-
torial wind time series (from the residual or an observ-
izational dataset) at a selected height in the lower
stratosphere. However, the choice of height would be
quite arbitrary (e.g., 40 hPa which is often favored). By
using the PCs, the element of arbitrary choice of rep-
resentative QBO height is eliminated. In addition, since
the time progression of the QBO is not independent of
altitude, the pattern of responses for a regression onto
a QBO time series in the lower stratosphere, such as 40
hPa, would not correctly attribute responses in the up-
per stratosphere. Hence, choosing to regress against
two QBO time series is a simple and more objective
method of representing the temperature (or zonal wind) QBO variations at differing altitudes. Taking the
second principal component is an obvious choice for
the second time series due to the orthogonal properties
of the principal components and the subsequent lack of
aliasing between the two indices in future regressions.
The QBO time series represented by the first and sec-
ond PCs are referred to as QBOa and QBOb, respec-
tively.

The length of the data record used was chosen so that
an exact number of QBO cycles was included in the
analysis to avoid introducing a linear trend due to an
inequality between the phase and/or magnitude of the
QBO at the bounds of the datasets. However, tests
showed that this precaution made negligible difference
to the regression patterns of the various signals, apart
from the linear trend, which is not examined here.

3. Results

The latitude–height structure of the solar, volcanic,
NAO, ENSO, and the QBO in the ERA-40 tempera-
tures and zonal wind are presented in this section. The
10.7-cm solar flux was normalized by 120 units of the
10.7-cm flux, the average difference between solar
minimum and solar maximum based on over 40 years of
observations. Thus the amplitude of the regression onto
the solar time series is the change per solar cycle
Similarly, both QBO time series were normalized by the average (QBO westerly minus QBO easterly) range so that the response patterns are the average temperature or zonal wind change over a QBO cycle. The volcanic, NAO, and ENSO terms were normalized by the standard deviation of their individual time series. Thus, the regression amplitude indicates the change resulting from a one standard deviation change in the index.

a. Solar response in the stratosphere

1) Annually averaged temperature response

Figure 2 shows the annual height–latitude structure of the solar signal in zonally averaged temperature (Kelvin per mean change in flux between solar maximum and solar minimum). The temperature response throughout the whole of the tropical stratosphere (above 16 km) is positive. This indicates increased temperatures during solar maximum years compared with solar minimum years. The pattern is statistically significant in the region 30°S–30°N and 35–48 km, peaking at an amplitude of over 1.75 K at 43 km (about 2.5 hPa). This agrees very well with the analysis of rocketsonde data by Keckhut et al. (2005). It is found to occur in all months (except February) with varying magnitudes and is likely to be predominantly a direct radiative effect via ozone absorption of increased shortwave radiation as predicted by models (see, e.g., Larkin et al. 2000). This is a robust feature of the solar response since a consistent temperature anomaly is present if only temperature data from one of the last two solar cycles (1979–90 or 1990–2001) are used in the regression analysis.

The pattern of response in Fig. 2 more closely corresponds to the result of S2000 than of H2004, especially the uninterrupted positive response throughout much of the tropical stratosphere. S2000’s statistically significant response over the equatorial latitudes between about 28 and 48 km, which maximizes between about 34 and 46 km, overlaps the maximum response exhibited in Fig. 2. Although the pattern of response is similar to that of S2000, the maximum amplitude of response is considerably larger, peaking at 1.75 K, compared with about 0.5–0.75 K; however, this value is still smaller (and at a lower altitude) than H2004’s maximum value of >2 K, irrespective of the difference in scaling.

Figure 2 also shows a large negative component in high northern latitudes at about 44 km (2 hPa) peaking at over −7.5 K at 55°N, resulting in a region of large anomalous meridional temperature gradient across midlatitudes. This contrasts with the study of H2004, which shows no such negative response in Northern Hemisphere (NH) midlatitudes. S2000, however, show at 70°N a negative response in the lower to mid stratosphere with a region of positive response above about 33 km, in general agreement with Fig. 2 below about 35 km. However, S2000 did not extend their analysis poleward of 70° (the latitudinal limit of the SSU data, although the MSU does extend to 90°), so we cannot compare the temperature response in high latitudes. These ERA-40 results are also in partial agreement with the rocketsonde study of Keckhut et al. (2005), subject to their error bars, who show a negative response at 20–54 km at northern midlatitudes (about 54°N). Within this height range they find a response of about −2.25 K at 43 km (for a 100 unit change in 10.7-cm radio flux), although they show a statistically significant maximum negative response of over −3 K at 30 km, which is not seen in Fig. 2.
Although the stratospheric response pattern in Fig. 2 is not entirely symmetric about the equator, there are large-scale coherent features seen in both hemispheres. A similar negative temperature response at high latitudes is also seen in the Southern Hemisphere (SH) polar region, maximizing at over 4 K at about 45 km, 90°S. The monthly averaged results (not shown) indicate that both these high-latitude features are predominantly winter phenomena and so are likely to be due to indirect dynamical effects.

Two significant lobes of positive temperature response are also present in the lower stratosphere in Fig. 2, each centered about 25° from the equator, with the NH lobe being slightly warmer than the SH lobe. They vary little between seasons, in terms of both positions and magnitudes. They are not present in the results of S2000 or H2004, possibly due to a lack of resolution in their datasets. Keckhut et al. (2005) find a similar significant positive response of ~1 K in their rocketsonde analysis centered at about 24 km in northern subtropics, slightly larger than the ERA-40 feature. In their analysis of NCEP–NCAR data between 1968 and 1999, Labitzke et al. (2002a) found regions of separate maxima in the subtropics (centered at about 30° latitude in each hemisphere) when temperature was correlated against 10.7-cm solar flux. In an earlier paper, van Loon and Labitzke (1999) found similar lobes to occur in at least (climatological mean) December and February if temperature data between 1968 and 1997 was grouped separately into QBO East years (determined by the direction of the zonal wind at about 45 hPa) before correlating against 10.7-cm solar flux, but not when grouped into QBO West years. They suggest that the influence of the 11-yr solar cycle on the stratosphere is dependent on the phase of the QBO.

We note also that the pattern of subtropical lobes with an equatorial minimum is rather similar to the meridional structure of the QBO (see Fig. 15) suggesting the possibility of a nonlinear interaction between the solar and the QBO signals that cannot be accounted for by this linear regression method. Aliasing between the QBO and the 10.7-cm flux is unlikely since these features remain, even if the solar flux time series is smoothed of all QBO time-scale variability.

As previously mentioned S2000 found a weak nonstatistically significant negative temperature response in the tropical lower stratosphere centered at about 22 km, which is not present in this current study. However, they did not account for volcanic activity, ENSO, or the QBO in their regression model and thus have not accounted for sources of atmospheric variability as completely as the present study, especially in the lower stratosphere. Lee and Smith (2003) examined a similar feature in a corresponding analysis of ozone data and found that it arose from volcanic and QBO influences. If the volcanic term is removed from our regression model, a weak, statistically nonsignificant, negative response appears in the ERA-40 data although the pattern of response elsewhere remains essentially unaltered. The same is true when the QBO is either completely neglected in the model or when it is imposed by specifying observed winds in the lower stratosphere instead of using the PC representation used in the analysis presented here (see section 2). Regressing over just 1979–98 has the same effect. The dominant statistically significant features of Fig. 2 remain, including the subtropical lobes, but a small negative feature appears over the equator at 24 km, which is not statistically significant. This emphasizes that caution should be taken when analyzing the possible solar response in the equatorial lower stratosphere since even small changes in the details of the regression model can lead to changes of the sign of the response in this highly variable region.

In summary, these results support the growing body of evidence that variability associated with the 11-yr solar cycle has a significant influence on stratospheric temperatures. However, there is still no consensus on the exact magnitude and spatial structure; longer and more consistent satellite observations are needed to resolve this issue. We note here that tests have shown that none of the discrepancies between the current work and that of S2000 and H2004 can be explained simply in terms of the slightly different lengths of the various datasets employed, nor the fact that H2004 used the Mg II index to represent solar variability rather than the 10.7-cm radio flux as was used in the current study and in S2000. We suggest that differences between the datasets employed is the primary reason for the large disagreement between the results of H2004 and those shown in the current analysis and in S2000. Although the exact reasons for the disagreement remain uncertain, H2004 notes that the SSU dataset compiled by NCEP–CPC and employed in his study may still exhibit significant offsets near the times of satellite transitions.

2) Annually averaged zonal wind response

The annual height–latitude structure of the solar signal in zonal wind fields is shown in Fig. 3. It is consistent (through thermal-wind balance) with the temperature response, which we have already discussed, with approximate hemispheric symmetry between 60°S and 60°N. The pattern of response is robust and the large-scale features outside certain equatorial regions within the stratosphere (see next paragraph) are insensitive to the details of exactly which terms are included in the regression model; when selected time series are excluded from the regression model, the variance that had previously been attributed to them is attributed to the noise, not to the other signals. The largest signal is in the subtropical upper stratosphere/lower mesosphere between 40°S and 40°N, centered around 50–55 km where the zonal winds are up to 6 m s⁻¹ more westerly in solar maximum than in solar minimum. A negative response is also present at high latitudes (>50°N) in both hemispheres although the amplitudes and patterns are not entirely symmetric about the equator. These
patterns of response are predominantly winter phenomena and will be discussed in more detail in the subsequent sections.

A region of statistically significant positive response is also found over the equator between about 33 and 36 km. It is surrounded by a pattern of nonsignificant negative response in the subtropics, which is again reminiscent of the pattern of QBO wind anomalies (see, e.g., Fig. 16). This may be an indication of a nonlinear solar cycle/QBO interaction in this region that cannot be isolated by a linear regression analysis such as this. Excluding the QBO terms from the multiple regression increases the amplitude, but reduces the significance of this feature. This could indicate that the QBO terms included here are not fully representing the QBO time variations in this region. Hence, perhaps there may be some QBO-like variability remaining that is being falsely attributed to solar variability, so the removal of the QBO terms would allow the possibility of increased aliasing of the QBO onto the solar signal, which consequently would result in the larger amplitudes. On the other hand, since removing the QBO terms from the regression decreased the confidence and increased the magnitude of the response, it could mean that, before they were excluded, the QBO terms were fitting the data well and, therefore, in effect removing the direct influence of the QBO in this region. As a consequence, this reduced the background variability, hence allowing a better solar fit to the data, thus resulting in the detection (with increased confidence) of what is perhaps a realistic and physical feature of solar response in zonal-mean zonal wind. This significant equatorial feature between about 33 and 36 km is also seen if the 10.7-cm radio flux is smoothed of QBO time-scale variability, thus implying that it is not merely a result of aliasing of the QBO by the solar time series. However, using this smoothed solar time series does change the amplitudes of the response in the tropical stratosphere (e.g., around 20–30 km) but not the direction of the responses, nor does it significantly improve the statistical significance of some of the previously nonsignificantly positive response amplitudes in this region.

In a further sensitivity test, QBO time series from the Berlin stratospheric data series of zonal winds (Labitzke et al. 2002b) were used to specify the QBO time series in place of the derived PCs. When either a single lower stratospheric value (40 hPa) or all available seven pressure levels between 70 and 10 hPa were used, it made negligible difference to the annual solar response in zonal-mean zonal winds apart from the feature over the equator around about 35 km, which again increased in magnitude but was reduced in significance.

In summary, the equatorial response at about 30–40 km is positive but there is some uncertainty concerning its size and significance. We suggest that this may be due to a nonlinear interaction with the QBO in this region, which cannot be accounted for in this analysis.

3) SEASONALLY AVERAGED RESPONSE

The temperature and zonal wind response associated with the 11-yr solar signal for individual seasons are shown in Fig. 4. In the temperature field, the feature common to all seasons is a positive signal over the equator at around 40–50 km. This is the region of high ozone mixing ratio, supporting the possibility that this signal is a direct thermal response to increasing TSI and UV heating in this region.

The dominant zonal wind response in the stratosphere is approximately consistent with the thermal response, with maxima above the regions of largest hori-
Horizontal temperature gradients. In both December–February (DJF) and June–August (JJA) there is a statistically significant positive response at 30°–50° latitude in the winter hemisphere, roughly coincident with the region of maximum winds forming the polar night jet. Since this signal is greatest in the winter hemisphere, this suggests a possible indirect (dynamical) response is at work in addition to any direct thermal response. Under solar maximum conditions the polar vortex is colder, stronger, and less disturbed than under solar minimum conditions. In fact, an examination of the response in the individual months that make up the DJF average (not shown) indicates that the vortex is stronger in solar maximum in both December and January (thus giving a strengthened vortex in the DJF average) but the reverse is true in February and March. This suggests that it may be the timing of the warmings that differ between solar maximum and solar minimum.

Fig. 4. Solar signal in (left) stratospheric temperature (K per solar maximum minus solar minimum) and (right) zonal wind (m s⁻¹ per solar maximum minus solar minimum): seasonally averaged, zonally averaged signal. Shaded areas denote statistical significance at the 5% (light shading) and the 1% (dark shading) levels.
This signature in the monthly data can be explained by warming events occurring preferentially in December/January in solar minimum years and preferentially in February (or later) in solar maximum years (see Gray et al. 2004). Since this signal is greatest in the winter hemisphere, this suggests a possible indirect (dynamical) response at work in addition to any direct thermal response.

The hemispheric asymmetry at high latitudes in the zonal wind annual average (see Fig. 3) throughout most of the stratosphere, and particularly the significant negative signal in the SH (>50°S), can be explained (in terms of its significance) by observing the disparities between the seasonal zonal wind responses, shown in Fig. 4. Notably, the SH response, >55°S, is invariably negative throughout all seasons and thus they combine to create the statistically significant negative response seen in the annual average. However, in the NH high latitudes (>50°N) no particular direction of response prevails across all seasons in the lower stratosphere, hence giving no statistically significant signal in the annual average response presented in Fig. 3.

In the lower-stratospheric temperature field the most notable statistically significant feature is the region of local positive response in the subtropics of each hemisphere during all seasons at around 20–25 km, already noted in the annual mean. The regions are distinct from each other, especially in DJF and March–May (MAM), with a region of reduced, or even negative (but not statistically significant) response over the equator. The maximum amplitudes are in the NH at around 1.5 K in DJF and in both hemispheres at around 1 K in JJA, when a relatively uniform significant signal is found to transect the whole tropical lower stratosphere. Although there are corresponding features in the zonal wind response, none of them are statistically significant apart from in JJA when the strengthened SH stratospheric polar vortex in solar maximum extends throughout the whole depth of the stratosphere and so includes the midlatitude lower stratosphere.

Figure 4 therefore confirms that the subtropical lower stratospheric signal is present all year round. This may be a direct thermal response, although the mechanism for a direct response at such low levels in the stratosphere is not known. It may also be an indirect response to the changes in the winter hemisphere dynamics that may change the mean meridional circulation. The sign of the signal can be explained in terms of reduced tropical upwelling during solar maximum compared with solar minimum. Any winter-induced anomaly from either hemisphere is likely to persist because the atmosphere in this region has a long radiative time scale (Scott and Haynes 1998). The fact that the structure of the signal is similar to the latitudinal structure of the QBO (see Fig. 15), which is the dominant feature of this part of the stratosphere, supports this possibility. Salby and Callaghan (2004) have noted a decadal-scale modulation of the QBO period in this region with the westerly phase displaying a factor 2 difference between solar minima and solar maxima. This will impact the induced vertical motion associated with the QBO and hence the observed temperatures. We also note that the Salby and Callaghan (2004) feature is an example of the frequency modulation of the QBO and is not a simple amplitude modulation. It is therefore a prime example of a nonlinear interaction between the solar cycle and the QBO signals that make it difficult to characterize the variability in this region.

b. Solar response in the troposphere

The annual temperature response is shown in Fig. 5, plotted on an linear pressure scale with a 0.1-K contour interval to emphasize tropospheric features, so that it may more easily be compared with previous results—for example, Haigh (2003). It is dominated by two partly significant vertical bands at midlatitudes. These extend down from around the tropopause through the whole vertical extent of the troposphere to the surface, maximizing at values in the region of 0.3–0.4 K in the NH and 0.2–0.3 K in the SH. The region of largest positive response in the NH is close to the surface and is significant at the 5% level between 25° and 60° and at the 1% level between 28° and 50°. The SH band of positive response also has a significant response at the surface. At the equator there is a region of significant negative response just below the tropopause. Both of these equatorial and subtropical features are in excellent agreement with the results of Haigh (2003).

Figure 6 shows the climatological distribution of the tropospheric zonally averaged zonal wind fields with the solar response superimposed. The results are very similar to the analysis of NCEP data by Haigh et al. (2004, manuscript submitted to J. Climate). The vertical banded structure is evident again. Haigh (1996) suggested that this vertical banded structure is associated with a latitudinal shift in the position of the tropospheric jets and, thus, the latitudinal extent of the Hadley cell. Our analysis supports this possibility. It is evident from this that the subtropical jet in the NH centered at 30°–40°N is slightly weaker under solar maximum conditions and that the midlatitude winds between 40° and 70°N are strengthened. A similar response is also evident in the Southern Hemisphere. The weakening of the background flow around the climatological subtropical jet accompanied by an increase in the anomalous zonal wind at higher latitudes implies that the latitudinal extent of the Hadley circulation is extended during solar maximum (relative to solar minimum), resulting in a slight poleward shift of the subtropical jets. This is in agreement with the modeling work of Haigh (1999) and Larkin et al. (2000). In addition, the midlatitude jet at 40°–50°S is strengthened and tightened, as indicated by the additional negative response poleward of 50°S.

The hemispheric difference between the zonal wind response at high latitudes in Fig. 6 can again be traced...
to differences in the seasonal response. In the SH the climatological SH jets are strong throughout the year (see Fig. 7) and the extratropical response is very similar in both position and sign in three out of four seasons (namely, two bands of negative response either side of a positive response at about 45°S). The exception is autumn (MAM) when the climatological wind field is slightly weaker and the response to solar forcing is considerably reduced and nowhere statistically significant (see Fig. 4). Hence, in the SH these consistent seasonal signals observed in the SH reinforce each other to give an annual averaged response that is similar in position if not in magnitude.

In comparison, the NH pattern of response exhibits more seasonal variability at high latitudes. For instance, the average DJF and September–November (SON)
patterns in Fig. 7 exhibit a vertical column of positive solar response between 45° and 60°N stretching from the surface into the stratosphere, while in MAM it is 20° farther north or south (or flipped in sign), and a negative response is now found between 45° and 60°N. In JJA the response in the NH, and specifically that between 45° and 60°, is rather ambiguous, which could be related to the relative weakness of the climatological background flow in this season. The latitudinal shifting of this region of positive response is a conceivable explanation for the nonsignificant positive response seen throughout much of the northern mid/high latitudes in the annual average (Fig. 6) and is likely to be related to the seasonal variability, in terms of both strength and position, of the tropospheric jets. It should be noted that in the SH each column of the banded structure has statistical significance at least at the 5% level except in MAM. However in the NH, apart from the central positive region in DJF, no region is statistically significant in any season at the 5% level. Thus, no clear seasonally dependent pattern of response is detected in this hemisphere due to the high degree of NH variability.

The GCM modeling study of Larkin et al. (2000) also shows a banded structure in both solstitial seasons (equinoctial seasons were not presented) with similar amplitudes to those in Fig. 7 for their solar maximum minus minimum zonal-mean zonal winds. As in Fig. 7 their maximum solar responses tend to be concentrated above the 400-hPa pressure surface. However, the vertical bands of response obtained from their study tend to be found in different latitudinal locations than those in the current analysis (although the exact spatial differences are sensitive to the ozone fields employed in their model). These differences may be indicative of a difference in the solar response, but as in the current data study, since the positions of the vertical bands are consistent relative to the positions of the jets, then the difference is more likely to be due to disparities between the ERA-40 climatology and that of the GCM employed in their study.

c. Volcanic response

The effect on ERA-40 zonal mean temperature of a one standard deviation change in stratospheric optical depth (from Sato et al. 1993) is presented in Fig. 8. Since the eruptions of El Chichón and Mt. Pinatubo in 1982 and 1991, respectively, both occurred around the same phase of the solar cycle (Fig. 1), there have been suggestions that some of the atmospheric behavior attributed to solar variability may actually be due to other factors, with volcanic activity being one of the causal effects (Salby and Shea 1991), and that aerosol effects could be confused with the 11-yr solar cycle effects in statistical tests (Solomon et al. 1996). Hence, even if there is a solar influence that is distinct from the volcanic influences, there nevertheless remains the problem of possible aliasing between the volcanic and solar terms in multivariate regression analysis, and thus mix-
ing between their derived signals (see, e.g., the comments of Haigh 2000 and Chanin et al. 1999).

A general inspection and comparison of Fig. 2 (solar) and Fig. 8 (volcanic) shows an unambiguous difference between the two patterns of temperature response. For example, the sign of the volcanic response is the same at all latitudes in the lower stratosphere, whereas the signature of the solar response is negative at high latitudes (see also Haigh 2003). Also, in the region of maximum positive solar-induced temperature response in the Tropics around 40–45 km, there is a weakly negative response in Fig. 8. Hence, although the timing of certain solar and volcanic events is similar, the temperature response to these forcings is quite different. We also note that excluding the solar term from the regression does not alter the pattern of response seen in Fig. 8.

In this study, a time series of stratospheric optical depths (Sato et al. 1993) was taken as one of the independent indices in the regression model in an attempt to separate the influence of volcanic activity from the solar signal. An alternative method would have been to reject specific years that correspond to periods shortly subsequent to a major volcanic eruption. This method of rejecting specific years in an attempt to deal with volcanic-induced effect on temperature was employed by Keckhut et al. (2005). Figure 9 shows the pattern of possible solar response that is produced when this approach is taken (1982, 1983 and 1991, 1992 were excluded). The main features of Fig. 2 are reproduced in Fig. 9 including the positive response in the upper equatorial stratosphere, the negative response at high latitudes, and also the subtropical lobes of positive response in the lower stratosphere. The tropospheric response is also similar; both show columns of heating in midlatitudes. However, there are differences, the most striking of which is the double peak of the maximum signal in the upper equatorial stratosphere and the two disjoint regions of statistically significant response centered at about 46 and 35 km. However, the maximum temperature response in this region remains at 1.75 K. The NH high latitude negative response loses much of its of statistical significance at the 5% level when the four years are excluded, but remains largely significant at the 10% level (not shown). In summary, the main features of the solar response are not dependent on the method employed to account for volcanic eruptions. Nevertheless, using the Sato et al. (1993) time series is the preferred method since it retains as many years as possible and avoids the task of arbitrarily selecting which years to exclude from the analysis.

Figure 8 indicates that the dominant statistically significant feature of the volcanic response is a positive response to increased optical depths, which peaks between 0.75 and 1 K at about 16–29 km. This feature is highly significant between ±40° latitude and agrees well with Haigh (2003). In the troposphere the signal is very small (in comparison to the stratosphere) and negative. The warming of the lower stratosphere and cooling in the troposphere ensuing from the injection of volcanic aerosols into the lower stratosphere is well known to be a result of both enhanced absorption and increased backscattering of solar radiation by sulphate aerosol in the lower stratosphere (see Robock 2000 and Eluszkiewicz et al. 1997) and of enhanced outgoing infrared radiation incident from the troposphere.

Figure 8 also shows cooling in the mid to upper stratosphere between about 30 and 50 km over the Tropics. The pattern of this cooling is reminiscent of the meridional structure of the QBO (see Fig. 15) and the SAO (semiannual oscillation), which is the dominant
pattern of variability in this height region. It may be evidence of aliasing of the volcanic signal by the QBO. However, it could also suggest a physical interaction of the volcanic signal with the QBO/SAO signal, possibly via a modification of the strength of the equatorial upwelling.

Above about 40–50 km there is a second warming region throughout the lower mesosphere at high latitudes. This vertical alteration of sign of the response is in reasonable agreement with that by Keckhut et al. (1995). The large, significant, temperature response at 80°S at 45–55 km peaks at ~2 K, more than twice the maximum in the tropical lower stratosphere. A similar signal is seen in the NH but is not statistically significant, probably due to increased variability in that region during winter. Examination of corresponding seasonally averaged plots (not shown) confirms that these polar signatures are primarily winter phenomena, suggesting that they are most likely to be an indirect response associated with mean meridional circulation changes.

The tropospheric temperature response is mostly negative, as would be expected, but is not homogeneous. There are three regions of high statistical significance, centered at about 40°S and about 30°N, reaching peaks of ~0.1 and ~0.2 K, respectively, with a final region peaking at value over ~0.35 K in the northern polar region. Except for a difference in arbitrary normalization of the stratospheric optical depth time series, this is consistent with the results of Haigh (2003).

The zonal wind response to volcanic forcing is seen in Fig. 10. The most notable feature is the highly significant response centered over the equator at 38 km. A similar but stronger feature appears when we only regress over 1991–94 inclusive, and no such feature appears when regressing over a time period without any notable volcanic activity, for example, 1997–2001. If the QBO terms are excluded from the regression analysis, this westerly anomaly increases considerably (by almost a factor of 2) but loses some statistical significance. The latter is expected since without the QBO term the “noise” variability in this region will be much higher. If the 40-hPa equatorial zonal wind time series (or a varying QBO time series depending on altitude) from the Berlin stratospheric data series is used in the regression analysis instead of the two principal components of the residual, as described in section 2, the significance is again reduced, as is the absolute value of the volcanic response. This is also to be expected because the observed zonal wind time series, being nonfiltered, are likely to contain a volcanic signal. Hence the QBO term, applied in this way, could falsely attribute some of the volcanic response to QBO variations as a result of degeneracy between the two terms. However, the qualitative pattern of response shown in Figs. 8 and 10 is robust to changes in the details of the regression model (viz., order of AR noise model or precisely which terms are included in the regression model).

The two regions of negative response above and below the 38-km feature in Fig. 10 are, for most parts, not statistically significant at the indicated significance levels. This signal could be due to weak degeneracy between the QBO terms and the volcanic term, or even artifacts of the assimilation. The former is difficult to avoid since both the eruptions of El Chichón and Mt. Pinatubo occurred at similar phases of the QBO cycle (westerly at 35 km). Alternatively, they could be an indirect response via increased heating of the tropical
lower stratosphere and, hence, increased upwelling, which could in turn modify the upper-stratospheric and mesospheric circulations (Chanin et al. 1999). The significant response between 20° and 50°N at an altitude of 25 km is also likely to be in response to circulation changes associated with anomalous upwelling due to the injection of volcanic aerosols.

The easterly zonal-mean zonal wind response around the tropical tropopause (about 20 km) extends poleward to ±50°–60° and extends downward into the troposphere in both hemispheres, producing easterly anomalies that reach the surface approximately centered around 50°N and 50°S. Between these two arms of anomalous easterlies, in the region 40°S–30°N, there are small westerly anomalies through most of the troposphere. If the tropospheric pattern is compared with the solar response pattern in Fig. 3 (or Fig. 6), it is evident that their patterns of response are dissimilar not only in the stratosphere but also in the troposphere. Although both figures show evidence of vertical banding, the latitudinal position of the negative bands is different. When the zonal wind response to volcanic forcing (Fig. 10) is compared to the climatological zonal wind fields (not shown), it is evident that there is no broadening of the Hadley circulation in response to volcanic forcing or associated poleward shift of the tropospheric jet, unlike in the solar response.

d. NAO and ENSO responses

The calculated temperature response associated with the NAO (Hurrell et al. (2001)) is illustrated in Fig. 11. Most of the signal is confined to the northern mid to high latitudes as expected. In addition to the dipole structure in latitude centered at 50°N in the troposphere there is also a clear temperature dipole in height in the northern polar region with a large positive response in the upper stratosphere (around 45 km) and a negative response beneath it, which is significant from 27 km down to the tropopause. It shows a maximum cooling of about 1.2 K (or 2 K for the DJF average) for a one standard deviation increase in the NAO index. This is much less than the 5 K quoted by Baldwin and Dunkerton (1999, hereafter BD99) in their linear regression of their Arctic Oscillation (AO) index and DJF-averaged zonal mean temperature from the NCEP-NCAR reanalysis. The possibility that this discrepancy is due to the different indices (NAO or AO index) used in the studies was checked by using the AO (obtained online at http://www.cpc.ncep.noaa.gov) in the multiple regression in place of the NAO index (not shown). Although the position of maximum cooling (about 2 K around 40 hPa at the North Pole) is similar to that given by BD99, the magnitude is still less. Further analysis of this difference is beyond the scope of this study, but other possibilities include the different methodology and datasets employed.

The dipole feature in Fig. 11 is primarily a winter response and is associated with significant changes in the strength of the polar night jet, as shown in Fig. 12, which shows the corresponding zonal wind response. It is therefore likely to be an indirect dynamical response, possibly due to changes in planetary wave propagation and subsequent changes in the mean meridional circulation. In agreement with BD99, the deep negative temperature response extends down from around 40 km into the upper troposphere and extends to the mid-latitudes with a slightly diminished magnitude. As was the case for Fig. 2 (solar) and Fig. 8 (volcanic), this tropospheric temperature response pattern presented
in Fig. 11 is in excellent agreement with Haigh (2003).

The zonal wind NAO pattern is strongest in NH winter (not shown), in agreement with other studies (BD99 and Thompson and Wallace 1998). Also in qualitative agreement with BD99, outside the winter season the signal is confined to the troposphere. The hemispheric asymmetry is a clear indication that the solar term and the NAO term are not being greatly aliased or confused. There is also a significant NAO negative response of $-0.75 \text{ m s}^{-1}$ at about $35^\circ$N, and a positive response of $1.25 \text{ m s}^{-1}$ at $50^\circ$–$60^\circ$N. These correspond to a weaker subtropical jet and a stronger midlatitude jet. This dipole extends vertically and has two separate maxima: one around 8 km and the other at 40 km. The latter is coincident with the polar night jet and corresponds to a stronger (more westerly) jet during the positive phase of the NAO.

Although the NAO index is determined by the NH tropospheric flow, Fig. 11 shows a small positive response in the southern polar regions of the lower mesosphere, which extends downward and equatorward into the stratosphere. Although not all of this tongue is statistically significant at the indicated levels, the area
of largest amplitude (about 65° S) is marginally significant, indicating a 0.6-K change in zonal mean temperature for a single standard deviation change in the NAO index. If real, this is most likely to be an indirect dynamical influence from the NH winter via changes in the mean meridional circulation.

The responses to ENSO (Battisti and Sarachik 1995; Harrison and Larkin 1998) are primarily tropical/subtropical in nature (Figs. 13 and 14), as expected. The ENSO is a highly zonally asymmetric phenomenon, so the response pattern shown here is likely to be diminished due to zonal averaging of the fields. Figure 13 shows that the tropical troposphere warms during ENSO, presumably due to an increase in latent heat release. This positive ENSO response in the tropical troposphere is present throughout the year but the largest response (0.5–1 K) occurs in the DJF average (not shown). There is also a cooling in the tropical lower and mid stratosphere, as noted by Reid (1994).

The corresponding zonal wind response (Fig. 14) shows a maximum tropospheric response at about 25°N and 25°S, in approximate thermal wind balance with the temperature fields. In the middle to upper stratosphere between 32 and 55 km there is an equatorial tripole feature, although only the lowermost segment is statistically significant. These low/midequatorial stratospheric features in both Figs. 13 and 14 are similar to the structure of the QBO, suggesting a possible interaction of ENSO and the QBO signals. Again, it could be evidence of aliasing of the QBO by ENSO in the regression model (see Fig. 1), a problem difficult to avoid. On the other hand, since an ENSO event is characterized by increased tropical heating and moist convection, it is likely that the accompanying change in tropical wave activity and vertical motion would influence the evolution of the QBO and the SAO. However, this cannot be determined via this study.

e. QBO response

Interannual variability of temperature and zonal wind in the tropical stratosphere is strongly influenced by the QBO (Baldwin et al. 2001). As described in section 2, the QBO variation in the dataset was isolated by first obtaining the first two principal components of the residual of a multiple regression of the ERA-40 stratospheric zonal mean zonal wind, fitted using an autoregressive noise model of order 3. These two PCs, which we refer to as QBOa and QBOb, account for more than 80% of the total variance (about 45% and 35%, respectively) of the equatorial stratospheric zonal wind field in the ERA-40 data between 1979 and 2001. Subsequently, the two PCs were then used in the full multiple regression analysis to characterize the QBO variation. In general, applying a PC-type analysis to a signal that travels uniformly without changing its shape gives two PCs that account for the majority of the variance. These PCs will be a quarter cycle apart and any orthogonal pair of these is valid. The QBO does not quite travel uniformly, but changes its shape as it goes through considerably varying environments. However, from Fig. 1 it appears that the PCs are close to being a quarter cycle apart, and hence, if desired, it would also be possible to form a linear combination of QBOa and QBOb to define the QBO. A similar method of obtaining QBO time series was performed by Randel et al. (1999) who used an optimal linear combination of temporally orthogonal QBO time series to isolate time variations that are coherent with the zonal wind QBO.

Figure 15 shows the two QBO zonal mean temperature signals from the regression model. By construction,
QBOa and QBOb are orthogonal. This can be inferred from Fig. 15, where for example, the region of maximum response over the equator at 35 km of QBOa (left plot) corresponds to a region of zero response in QBOb.

A three-cell pattern in height (and also in latitude) with alternating sign is evident in both figures. Previous work (e.g., Randel et al. 1999; Baldwin et al. 2001) has shown the QBO to have only a two-cell structure in height, with maxima only in the lower and middle stratosphere, but Fig. 15 clearly shows a third anomaly in the upper stratosphere, probably as a result of improved satellite data assimilation in the ERA-40 dataset. A comprehensive description of the QBO signal in the ERA-40 dataset is provided by Pascoe et al. (2005).

In addition to the maximum signal over the equator, there are smaller QBO temperature components present at higher latitudes (20°–50°) of each hemisphere associated with the QBO-induced meridional circulation. These features are approximately symmetrical in the annual average and are out of phase with the main equatorial QBO temperature signal. However, as noted by Randel et al. (1999), these mid-latitude signals are not present all year round. They occur primarily during winter in the respective hemi-
sphere and could be linked to the seasonal synchronization of the Brewer–Dobson circulation.

Figure 16 shows the corresponding zonal wind fields. The “horseshoelike” pattern in the tropospheric zonal wind response of the QBOb pattern is similar to the negative response in zonal wind due to increases in solar irradiance (see Fig. 3). Haigh (1996) suggested that the horseshoe pattern in zonal wind response to the 11-yr solar cycle was due primarily to increased heating in the lower equatorial stratosphere during solar maximum conditions, which increased the temperature gradient and hence the latitudinal extent of the Hadley circulation. It is difficult to determine whether these similar tropospheric negative zonal wind responses in the solar and QBOb patterns have a similar origin, given the difference in latitudinal and height structures of the solar and QBO responses. However, it is clear that they occur in regions of strong latitudinal and vertical gradients, so only a small but consistent shift would be required to create them. Interestingly, the QBOa zonal wind response in Fig. 16 has a small positive zonal wind maximum near the tropical tropopause, which is the opposite sign to the QBOb pattern and extends only weakly downward to the surface. This suggests a sensitivity of the strength and sign of the tropospheric horseshoe pattern response to the height of the anomalies in the lower stratosphere or, specifically, the nature of the QBO time series employed in the regression. A separate linear regression analysis has been used to separate the climate patterns associated with a number of different atmospheric variations (solar, volcanic, NAO, ENSO, QBO). The patterns of response amplitudes in both zonal-mean zonal wind and temperature for each of these have been distinguished and have been shown to be distinct. The patterns presented have been tested by including and excluding various terms and by using differing order of AR models in the regression and found to be robust in terms of both amplitude and statistical significance in all regions, with the possible exception of the equatorial stratosphere where, in some circumstances, aliasing between signals and the QBO is unavoidable. Degeneracy in the overall regression model is at a negligible level, hence prompting confidence in the analysis. In particular the solar and the volcanic signal are completely separable as their individual patterns of large-scale response are generally invariant to the inclusion or exclusion of the forcing time series of the other.

Throughout the low latitudes (30°S–30°N) the stratosphere (16–55 km) warms in response to the 11-yr solar cycle in the ERA-40 data (Fig. 2). A large region of highly statistically significant positive response is found over the equator between about 35–50 km, peaking at about 43 km with an amplitude of 1.75 K. This warming is present in all seasons and is therefore likely to be a direct radiative response to solar irradiance and UV absorption by ozone in this region during solar maxima.

4. Summary and discussion

A linear multiple regression analysis has been used to separate the climate patterns associated with a number of different atmospheric variations (solar, volcanic, NAO, ENSO, QBO). The patterns of response amplitudes in both zonal-mean zonal wind and temperature for each of these have been distinguished and have been shown to be distinct. The patterns presented have been tested by including and excluding various terms and by using differing order of AR models in the regression and found to be robust in terms of both amplitude and statistical significance in all regions, with the possible exception of the equatorial stratosphere where, in some circumstances, aliasing between signals and the QBO is unavoidable. Degeneracy in the overall regression model is at a negligible level, hence prompting confidence in the analysis. In particular the solar and the volcanic signal are completely separable as their individual patterns of large-scale response are generally invariant to the inclusion or exclusion of the forcing time series of the other.

Throughout the low latitudes (30°S–30°N) the stratosphere (16–55 km) warms in response to the 11-yr solar cycle in the ERA-40 data (Fig. 2). A large region of highly statistically significant positive response is found over the equator between about 35–50 km, peaking at about 43 km with an amplitude of 1.75 K. This warming is present in all seasons and is therefore likely to be a direct radiative response to solar irradiance and UV absorption by ozone in this region during solar maxima.
The location of this region of positive equatorial response is in general agreement with the Keckhut et al. (2005) study of rocketsonde data and with S2000’s linear regression analysis of SSU/MSU irradiance data, but does not agree as well with H2004’s analysis of NCEP–CPC data. However, the amplitude of the temperature response is larger than S2000 and is closer in magnitude to that of H2004, who found a maximum statistically significant) positive response of over 2 K at 48 km. H2004 also found a significant negative response of over −1 K between about 26 and 35 km at low latitudes, which is not reproduced in this ERA–40 analysis. These differences may be due to the different datasets employed but may also arise from differences in the regression analysis performed.

A negative temperature response to the solar cycle is found at high latitudes of both hemispheres. A similar (but not significant) negative response is present in S2000’s analysis, although their region of negative response does not extend into the upper stratosphere. The reason for this discrepancy is unclear. Keckhut et al. (2005) also find a statistically significant negative response at midlatitudes throughout the stratosphere in their localized (i.e., nonzonal mean) rocketsonde analysis but this peaks at 30 km, a much lower altitude than seen in ERA–40. The anomalous meridional temperature gradient gives rise to a strong solar-induced zonal wind response in the upper stratosphere/lower mesosphere of both hemispheres (Fig. 3). These midlatitude features are predominantly a wintertime phenomenon associated with the strength of the polar night jet (Fig. 4). They are therefore likely to be due to indirect (dynamical) effects. Analysis of monthly responses suggests that it is the timing of the stratospheric winter warmings that is influenced by increases in solar cycle activity (Gray et al. 2004).

In the lower stratosphere the most notable, statistically significant, temperature responses to the 11-yr solar cycle are regions of local positive maxima (0.5–1.5 K) in the subtropics of each hemisphere during all solar cycle are regions of local positive maxima (0.5–1.5 K) in the subtropics of each hemisphere during all seasons at around 20–25 km and 20°–25° (Fig. 2). Although this may be a direct thermal response, there is currently no known mechanism for one at such altitudes. The signal may be explained in terms of an indirect response to changes in winter hemisphere dynamics via the meridional circulation and, hence, modulation of the tropical upwelling. Their latitudinal structure is very similar to the latitudinal structure of the QBO (see Fig. 15), which is the dominant feature of this part of the stratosphere. This suggests that they may be associated with nonlinear interaction between the solar and QBO signals, but this cannot be confirmed by the linear regression technique employed in this study.

In the troposphere the annual temperature response to the 11-yr solar cycle shows two vertical columns of statistically significant positive response at midlatitudes in both hemispheres. The associated zonal wind response shows a “horseshoe” pattern with a vertical banded structure indicating a weakening of the subtropical tropospheric jets in solar maximum coupled with a change in their position (Figs. 3 and 6). Haigh (1996) suggested that this horseshoe pattern in zonal wind response to the 11-yr solar cycle is due primarily to increased heating in the lower equatorial stratosphere. Our results are consistent with Haigh’s hypothesis, although the amplitude of the response suggests that it may be sensitive to the latitudinal and/or height structure of the anomalous heating.

In summary, these results support the growing body of evidence that the 11-yr cycle of solar variability has a significant influence on the atmosphere–climate system. However, the exact magnitude and spatial structure of this effect is still unclear; longer and more consistent satellite observations are needed to resolve this issue.

Acknowledgments. The authors wish to thank Myles Allen, David Andrews, William Ingram, and Gerald North for their helpful comments and advice.

REFERENCES


