The Influence of the Solar Cycle and QBO on the Late-Winter Stratospheric Polar Vortex

CHARLES D. CAMP AND KA-KIT TUNG

Department of Applied Mathematics, University of Washington, Seattle, Washington

(Manuscript received 31 August 2005, in final form 24 July 2006)

ABSTRACT

A statistical analysis of 51 years of NCEP–NCAR reanalysis data is conducted to isolate the separate effects of the 11-yr solar cycle (SC) and the equatorial quasi-biennial oscillation (QBO) on the Northern Hemisphere (NH) stratosphere in late winter (February–March). In a four-group [SC maximum (SC-max) versus minimum (SC-min) and east-phase versus west-phase QBO] linear discriminant analysis, the state of the westerly phase QBO (wQBO) during SC-min emerges as a distinct least-perturbed (and coldest) state of the stratospheric polar vortex, statistically well separated from the other perturbed states. Relative to this least-perturbed state, the SC-max and easterly QBO (eQBO) each independently provides perturbation and warming as does the combined perturbation of the SC-max–eQBO. All of these results (except the eQBO perturbation) are significant at the 95% confidence level as confirmed by Monte Carlo tests; the eQBO perturbation is marginally significant at the 90% level. This observational result suggests a conceptual change in understanding the interaction between solar cycle and QBO influences: while previous results imply a more substantial interaction, even to the extent that the warming due to SC-max is reversed to cooling by the eQBO, results suggest that the SC-max and eQBO separately warm the polar stratosphere from the least-perturbed state. While previous authors emphasize the importance of segregating the data according to the phase of the QBO, here the same polar warming by the solar cycle is found regardless of the phase of the QBO.

The polar temperature is positively correlated with the SC, with a statistically significant zonal mean warming of approximately 4.6 K in the 10–50-hPa layer in the mean and 7.2 K from peak to peak. This magnitude of the warming in winter is too large to be explainable by UV radiation alone. The evidence seems to suggest that the polar warming in NH late winter during SC-max is due to the occurrence of sudden stratospheric warmings (SSWs), as noted previously by other authors. This hypothesis is circumstantially substantiated here by the similarity between the meridional pattern and timing of the warming and cooling observed during the SC-max and the known pattern and timing of SSWs, which has the form of large warming over the pole and small cooling over the midlatitudes during mid- and late winter. The eQBO is also known to precondition the polar vortex for the onset of SSWs, and it has been pointed out by previous authors that SSWs can occur during eQBO at all stages of the solar cycle. The additional perturbation due to SC-max does not double the frequency of occurrence of SSWs induced by the eQBO. This explains why the SC-max/eQBO years are not statistically warmer than either the SC-max/wQBO or SC minimum/eQBO years. The difference between two perturbed (warm) states (e.g., SC-max/eQBO versus SC-min/eQBO or SC-max/eQBO versus SC-max/wQBO), is small (about 0.3–0.4 K) and not statistically significant. It is this small difference between perturbed states, both warmer than the least-perturbed state, that in the past has been interpreted either as a reversal of SC-induced warming or as a reversal of QBO-induced warming.

1. Introduction

In the winter polar stratosphere, the solar cycle (SC) signal was thought to be detectable only when the data are stratified according to the phase of the equatorial quasi-biennial oscillation (QBO; Labitzke 1987; Labitzke and van Loon 1988, hereafter LvL88). Why this is so has never been explained, and this has created confusion as to the mechanism of the solar cycle response in the stratosphere. Is the QBO essential for the amplification of the solar cycle signal? What is the dynamical mechanism linking the two phenomena? A prevailing school of thought [see recent discussions in

Corresponding author address: Charles D. Camp, Dept. of Applied Mathematics, University of Washington, Box 352420, Seattle, WA 98195.
E-mail: cdcamp@amath.washington.edu

DOI: 10.1175/JAS3883.1

Labitzke (2003) and Labitzke (2004)] believes that the structure of the solar cycle response takes different forms according to the phase of the QBO. A different possibility, which we shall explore here, is that the solar cycle influence is not intricately tied to the QBO mechanism. The solar cycle merely represents an "additional" perturbation, comparable in magnitude to the QBO, to the least-perturbed state of the polar stratosphere, which is represented by the state of solar minima (denoted by SC-min) and westerly QBO (denoted by wQBO).[1]

The variation of total solar irradiance over an 11-yr solar cycle is known to be small, having been measured by satellites at approximately 0.1% of the solar constant (Willson et al. 1986; Lean 1991). It is generally accepted that a dynamical amplifier to the radiative forcing is needed to account for the observed solar cycle signal in the lower atmosphere (the lower stratosphere and the troposphere). The variability in the solar ultraviolet wavelengths is larger, at approximately a few percent. Energy at these wavelengths is absorbed by ozone, which is abundant in the stratosphere. It follows then that the atmosphere's solar cycle response should be largest over the lower latitudes in the stratosphere where the solar radiation is strong. Both models (e.g., Haigh 1996) and observation (Labitzke 2001; Haigh 2003; Hood and Soukharev 2003; Crooks and Gray 2005) found a tropical solar cycle signal of about 1.5–2 K in the upper stratosphere and 0.5–1 K in the lower stratosphere. Yet the largest signal is found during winter over the pole (Labitzke 2001), where the solar radiation is the least. The magnitude of the solar cycle warming during the polar night is about 7 K, from solar minimum to solar maximum (see later in this paper), which is much larger than that observed over the Tropics and larger than can be explained by radiative consideration alone.

Over the polar stratosphere during winter, the interaction of the QBO and solar cycle superficially appears to be complicated and not yet understood. One puzzling prior result we wish to reexamine first is that pertaining to the reversal of the SC-max warming by the easterly QBO (denoted by eQBO). Labitzke (1987) found that during wQBO years, defined by the 50-hPa equatorial winds, the 30-hPa North Polar temperature averaged over 4 months (November–February) is positively correlated with the solar flux, being warmer during the SC–max than during the SC-min, with a correlation coefficient of 0.78 and a statistical confidence level of 99.9% in a Student's t test. In the eQBO years, the correlation coefficient was found to be -0.32 and not statistically significant. She further noted that in the eQBO years stratospheric sudden warmings (SSWs) occur during all stages of the solar cycle, but that in the wQBO years no SSWs were found when the sunspot number was below 110. This is consistent with her earlier report (Labitzke 1982) that more midwinter SSWs occurred in the eQBO years, while in wQBO years midwinter SSWs occurred only during sunspot maxima. Later, LvL88 found that the correlation of the unstratified January–February mean North Polar temperature at 30 hPa and the solar flux for the 32 winters in the period 1956–87 was insignificant at a correlation coefficient of 0.14. However, when stratified, the correlation was a positive 0.76 for the wQBO years, and -0.45 for the eQBO years. Both *positive* and *negative* correlations were now found by the authors to be statistically significant using a Monte Carlo simulation. This oft-cited Monte Carlo test was their most rigorous statistical test yet, and it appears to substantiate the notion that QBO interacts with the solar cycle in an intricate way so that the warming one normally expects to occur during SC-max is reversed to cooling when the years coincide with eQBO years. We have reexamined the Monte Carlo simulation of LvL88 and reproduced their Fig. 2 in our Fig. 1. For the period considered by them, 1956–87, the observed correlation of the 30-hPa January–March mean temperature at the North Pole and the solar flux was found to be $r_w = 0.76$ during the westerly phase of the QBO and $r_e = -0.43$ during the easterly phase of the QBO. As shown in the figure (and consistent with LvL88), the observed pair of correlations is so far from the point of no correlation that very few randomly generated time series can achieve a greater correlation pair. A circle of radius $\sqrt{r_w^2 + r_e^2}$ was drawn by LvL88, who pointed out that in the simulation of 10 000 series only 25 have correlations outside the circle and concluded that "this is convincing evidence that our original correlations of 0.76 and -0.45 did not occur by chance." However, the length of the radius of the circle is determined primarily by the strength of the correlation during the westerly phase. To examine the separate correlations during each phase of the QBO, straight lines (as drawn in our figure) should be used instead of circles. While there are very few randomly generated points achieving an r_w magnitude greater than the observed value, there are many points with an r_e magnitude greater than the observed value. Counting these points reveals that, while the ob-

[1] We think the least-perturbed state should actually be the SC-min/wQBO and cold El Niño, and the largely independent perturbations to this state are the eQBO, solar maximum, and warm El Niño; see Camp and Tung (2007). However, the length of the data record is not long enough for us to consider the separation of the eight groups.

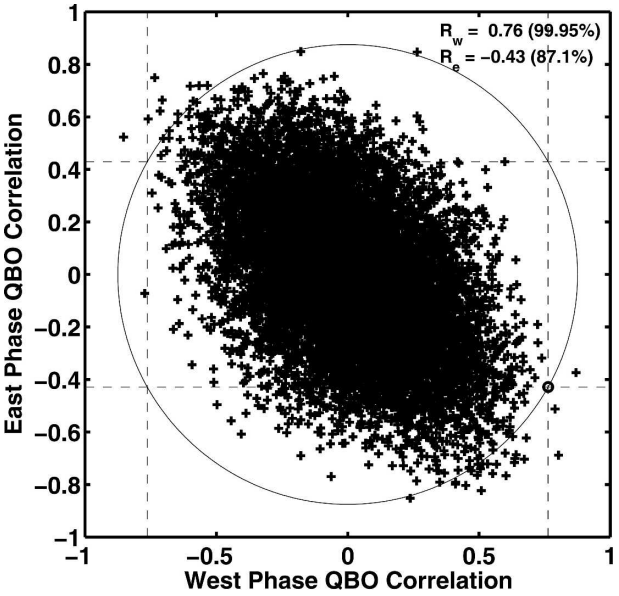

FIG. 1. Reconstruction of LvL88's Fig. 2. Paired correlations between the mean 30-hPa North Pole temperature and the 10.7-cm solar flux when partitioned by the phase of the 45-hPa tropical QBO index. Mean values for January–February from 1956 to 1978 are used. Statistical significances for the observed correlations determined by calculating the paired correlations from partitioning 10 000 first-order autoregressive surrogate time series ($p_1 = -0.34$).

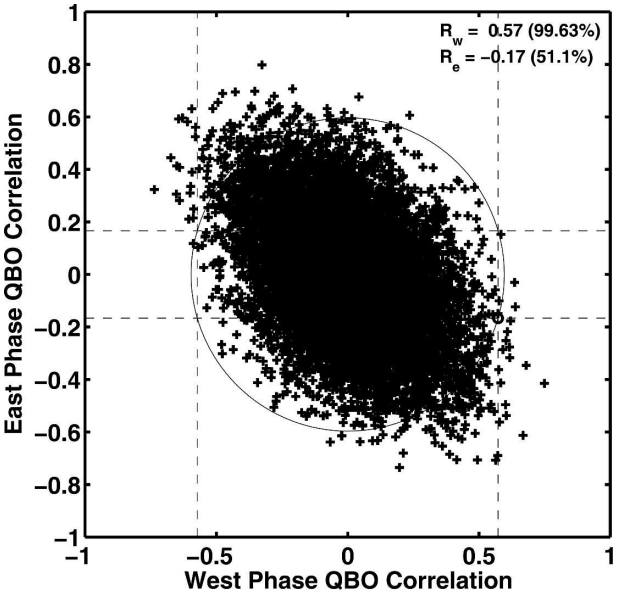

FIG. 2. As in Fig. 1 but for data from 1956 to 2001 ($p_1 = -0.33$).

served correlation with the solar cycle during the westerly QBO is statistically significant at 99.9% confidence level, the correlation is not significant during the easterly QBO. Extending the length of data by 15 more years further lowers the statistical significance of the easterly correlation with the solar cycle, to 51%, while the westerly correlation remains statistically significant at 99.6% (see our Fig. 2). The "bootstrap" Monte Carlo test employed by LvL88 in their Fig. 1 appears to suffer a different problem. The statistical test was done after the data were segregated into the eQBO and wQBO years, yielding two decadal signals, which were then correlated with the solar flux—also a decadal signal. To test the statistical significance of the correlations of these decadal signals, synthetic time series were created by a bootstrap method. Many of these constructed time series do not have any decadal cycle and therefore have no potential of being correlated with the 11-yr solar cycle. As a consequence, the statistical significance of the observed correlation is inflated by the presence of these "nonisospectral" time series. In a recent update, Labitzke (2005) concluded that "the 30-hPa North Pole temperatures are positively correlated with the 11-year sunspot cycle during the west phase of the QBO while no correlation exists for the east phase of the QBO directly over the North Pole." We concur with this latest conclusion.

One explanation of the negative correlation in the results for Labitzke (1987) and LvL88 for the North Polar temperature may be that, while the region of positive correlation in the wQBO years is geographically wide, the region of negative correlation in the eQBO years is a small area near the North Pole. The use of the temperature at a point (the North Pole) may have exaggerated a regional effect. In our work, we seek a coherent meridional pattern over the Northern Hemisphere that is correlated with the solar cycle, thus minimizing the regional effects. Another explanation is that the warming is measured relative to a reference state. If that reference state is a perturbed state that is already warm, such as the mean during eQBO years, the difference may become negative. However, the difference is very small and not statistically significant. Any information about the magnitude of the difference is lost when a correlation coefficient is used, as was previously done.

There have also been suggestions that the solar cycle–QBO interaction may be artificially created by undersampling when data are stratified according to the phases of QBO (the "unphysical explanation"). As pointed out by Salby and Shea (1991) and Salby et al. (1997), the symptoms of such a phenomenon are (i) the 11-yr cycle is not apparent in the unstratified time series, (ii) but appears when stratified according to even or odd years or according to the phase of the QBO, and (iii) the 11-yr cycle seen in one stratum (e.g., the westerly phase of the QBO) is opposite in amplitude to the 11-yr cycle seen in the other stratum (e.g., the easterly

phase of the QBO). All three of these symptoms apparently fit the results reported by Labitzke (1987) and LvL88. This is the reason why the oppositely signed correlations cause concern. The results to be reported here, however, do not share these symptoms, nor do the results of LvL88 when reinterpreted as is done above.

The mechanism of QBO's interaction with the polar stratosphere is better understood, and yet many of the reported observational results have not been explained. Holton and Tan (1980, 1982) discovered what is now called the Holton–Tan effect. They found that in composites, according to the phase of the equatorial QBO at 50 hPa, the polar winter temperature is warmer and the polar vortex is more perturbed by planetary waves when the equatorial QBO is in its easterly phase than in its westerly phase. For a possible mechanism, they cited the work of Tung and Lindzen (1979a) on the effect of the position of the zero-wind line on the stationary planetary waves: As the zero-wind line moves more poleward during eQBO, the westerly waveguide for stationary waves is narrowed and located more poleward. This tends to focus the planetary waves more poleward, making the polar vortex more disturbed and hence warmer. Our understanding of wave–mean flow interaction in the stratosphere has progressed considerably since the late 1970s. Instead of the quasi-linear picture of wave–mean interaction in which the zero-wind line plays a crucial role, it is now understood that the dynamics is highly nonlinear in the stratosphere, and planetary waves break in surf zones (McIntyre and Palmer 1984). Nevertheless, the planetary waves do tend to break more poleward during the QBO easterly phase than during the westerly phase when the waveguide is wider and the wave flux is directed more equatorward. Holton and Tan (1980) divided the winter into early winter (November–December) and late winter (January–March). They found, using 16 years of data, that in early winter wavenumber-1 amplitude at 50 hPa is about 40% greater in eQBO than in wQBO, at the 99% confidence level in a Student's t test. This positive result was later questioned by Naito and Hirota (1997, hereafter NH97) who found that the Holton–Tan result for 1962/63–1977/78 failed to hold in the longer record up to 1993/94. NH97 conjectured that the difference found by Holton and Tan was due to the solar cycle because the period used by them happens to contain two solar minima and one maximum. They therefore suggested that Holton and Tan's result could not be valid in general, it being applicable only to periods with more solar minima than maxima. Gray et al. (2001)'s findings echo these: For the shorter 26-yr period of 1964–90, with a bias toward the solar minimum, the correlation between January–February North Polar

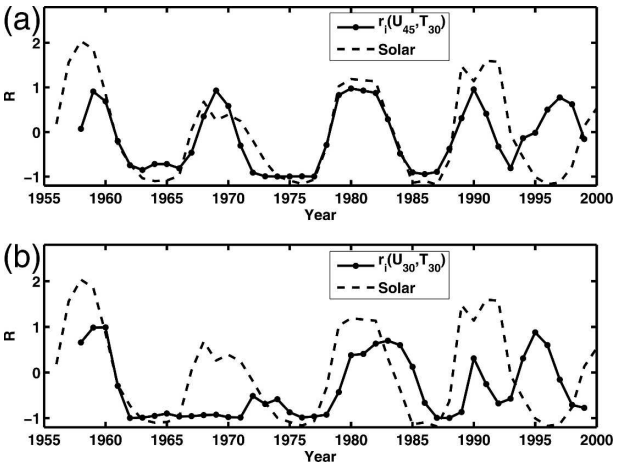

FIG. 3. (a) Three-year running correlation between the 45-hPa equatorial zonal wind in January and the January–February average of the 30-hPa North Pole temperature. The dashed line is the normalized 10.7-cm solar flux, December–February average. (b) As in (a) but using the 30-hPa equatorial zonal wind.

temperature in the lower stratosphere with the equatorial QBO wind is −0.4 (the negative sign meaning that the pole is warmer during easterly QBO years), but reduces to −0.25 for the 44 winters from 1955 to 1999, including four full solar cycles. Holton and Tan (1980) also found that during late winter the behavior of wavenumber 2 was unexpectedly opposite, being about 60% stronger, in the composite mean, during the wQBO than in the eQBO at 96% confidence level. However, when four more years were added to the sample, they found that the significance level dropped to about 90%.

Using unstratified data, Kodera (1993, hereafter K93) computed a running correlation between the equatorial zonal wind at 45 hPa and the 30-hPa North Polar temperature from Freie Universitat Berlin (FUB) for the period 1956–91. The resulting correlation coefficient is positive during the solar maximum and negative during the solar minimum, suggesting that the solar cycle result uncovered by LvL88 may be real despite the possibility of aliasing. Salby and Callaghan (2004) provided a Monte Carlo test of a similar running time-mean correlation with the solar flux and found it to be statistically significant. There is still no explanation of the puzzling positive correlation with the QBO during the solar maximum, which was deemed "inconsistent" with Holton–Tan (Salby and Callaghan 2004). We have reexamined this result by repeating K93's calculation for the longer time period of 1954–2001, the full extent of the FUB stratospheric temperature used therein. Figure 3a shows that the K93 result holds for three solar cycles but fails during the solar minimum of the late 1990s. Furthermore, when a different level, 30 hPa, of

the equatorial wind is used in place of the 45-hPa wind used in K93 as an index for the phase of the QBO, there is no reversal during the solar maxima circa 1970 (see Fig. 3b). A possible explanation may have already been provided by the work of Salby and Callaghan (2004): During each of the solar maxima prior to 2000, the phase of the equatorial QBO changed during the winter season. This implies that the choice of vertical level defining the QBO phase becomes critical. Salby and Callaghan (2004) further state that during these SC–max years, the wintertime-mean equatorial QBO, integrated upward over 30–10 hPa, actually reverses. K93 used a lower-level equatorial wind, 45 hPa, to correlate with the 30-hPa North Pole temperature. When viewed this way, the so-called reversal of the Holton–Tan mechanism during the SC–max does not appear so mysterious anymore; it may just be a matter of deciding, from a dynamical perspective, which level of the equatorial wind the planetary waves "see." There is, as yet, no consensus on the location of this level. Using the 40-yr European Centre for Medium-Range Weather Forecasts (ECMWF) Re-Analysis (ERA-40) data, which extends to 0.1 hPa, Pascoe et al. (2005) discovered a threefold structure in height for the QBO wind at the equator, complicating the determination of the phase of the QBO even further.

Stratifying the period 1955–99 according to whether the solar flux is above or below the mean for the period, Gray et al. (2001) found that the January–February mean North Polar temperature in the lower stratosphere is correlated with the phase of equatorial QBO with a correlation coefficient of -0.7 when only solar minimum years are considered. When only solar maximum years were used, the correlation is a much weaker 0.1. The fact that the coefficient here is positive, implying a reversed Holton–Tan QBO effect, is probably immaterial because the correlation is not statistically significant for the solar maximum years. It is perhaps more appropriate to say, as suggested by Gray et al. (2004), that the Holton–Tan effect has been "disrupted" by the solar maximum, rather than to suggest that the Holton–Tan effect is "reversed."

The above paragraphs provide a brief review of the perturbations to the polar stratosphere along two perspectives: the solar cycle's effect on the polar stratospheric temperature, possibly obscured by the eQBO, and the QBO's effect on the same temperature, possibly obscured by the solar maximum. The combined perturbation of SC-max/eQBO is yet to be clarified. We will show that the problem is related to the identification of a reference state. The picture becomes much clearer and the results more statistically significant when "warming" is defined as relative to an unperturbed or a least-perturbed state. In our data analysis, we shall first consider all of these effects simultaneously without prejudice. In a four-group analysis of SC-min/wQBO, SC-min/eQBO, SC-max/wQBO, and SC-max/eQBO we will identify the first state as being a least-perturbed (cold pole) state that is statistically distinct from the rest, and the other states as providing similar warming perturbations to this least-perturbed state. This is consistent with Fig. 14 of Gray et al. (2004), which visually shows less variability in the North Polar temperature during the SC-min/wQBO. We now provide a rigorous statistical confirmation. We will then consider separately the effects of the solar cycle and QBO perturbations, and show that the spatial pattern of both perturbations is similar to the meridional structure that one would expect from SSWs. We hope that our observational analysis can provide the circumstantial evidence of SSW being the dynamical amplifier for the solar cycle signal in the polar stratosphere. We further obtain a statistically significant result that the combined SC-max/eQBO perturbation warms the polar vortex relative to the least-perturbed state. There is no reversal of either SC-induced or QBO-induced warming when viewed in this way.

In this work, we emphasize rigorous (conservative) statistical tests of our results. Although a confidence level as low as 75% has sometimes been adopted as significant in solar cycle research, we shall use the higher threshold of the 95% confidence level for statistical significance. We call a result "barely significant" if the confidence level falls between 90% and 95%. The result is deemed "statistically not significant" if the confidence level falls below 90%.

2. Methodology

Linear discriminant analysis (LDA) is a long-standing statistical technique used for the classification of multivariate data into predefined groups (Wilks 1995; Ripley 1996). Schneider and Held (2001) have demonstrated its usefulness for identifying spatial patterns associated with interdecadal variations of surface temperatures. This application of LDA focuses on coherent spatial variability as opposed to total variance at each individual grid point. As such, it can often isolate large-scale spatial variability with greater statistical significance than studies using correlations of individual time series. On the other hand, the method may not find statistically significant differences among groups if there are no coherent spatial differences. Compared to multiple-regression methods applied to time series at each location independently, the present LDA technique takes advantage of the spatial information to de-

termine the optimal spatial pattern or patterns that best distinguish the behavior between the different groups of observations. LDA also improves upon traditional spatial filter techniques, such as composite-mean differences, in that it can simultaneously analyze more than one underlying process (two groups) and in that it extracts a much cleaner temporal signature of the desired behavior(s). In this study, we apply the technique to identify the spatial patterns that best distinguish the behavior of the stratosphere during the westerly and easterly phases of the equatorial QBO and during the SC-max and SC-min.

For a centered (zero time mean) set of multivariate observations, $X(t, \mathbf{x})$ we seek the pattern, denoted by $P(\mathbf{x})$, which best distinguishes between the groups of observations (such as SC-max/wQBO years versus SC-min/wQBO years). The optimization procedure searches for a spatial pattern normalized by the variances and covariances of the variables. This *scaled* spatial pattern is denoted by $u(\mathbf{x})$, where $P(\mathbf{x}) = X^{\mathrm{T}}X\,u(\mathbf{x})$. The elements of $u(\mathbf{x})$ then represent the relative importance of a given location to a chosen separability measure, denoted by $\mathcal{R}$. For an arbitrary $u(\mathbf{x})$, we can define an $\mathcal{R}$ in the following manner. Each year's observation from the centered dataset is assigned to one of the predefined groups and then each group of observations is projected onto $u(\mathbf{x})$. The projected variance of the entire dataset can be partitioned into a *between-group* variance and a *within-group* variance. The between-group variance captures variability associated with the preselected process or processes (e.g., SC and/or QBO), while the within-group variance represents the variability that still exists within each group associated with other phenomena (e.g., ENSO). We wish to maximize the former and minimize the later; therefore, we define a suitable separability measure, $\mathcal{R}$, as the ratio of between-group variance to within-group variance. The desired scaled pattern, $u(\mathbf{x})$, is the vector that maximizes $\mathcal{R}$. When the original centered dataset is projected onto $u(\mathbf{x})$, we get the first canonical variate, $C(t) = Xu(\mathbf{x})$, a time series whose elements are "scores" for each observation. The associated spatial pattern, $P(\mathbf{x})$, can now be recovered by regressing the data onto $C(t)$, that is, $X(t, x) = C(t)P^{\mathrm{T}}(x) + \epsilon(t, x)$. (See appendix A for further details.) In other words, $P(\mathbf{x})$ is the *spatial pattern* that best distinguishes between the *groups of observations* while the time series $C(t)$ represents an *"index"* for *that spatial pattern.*

With only two groups (i.e., SC-max/wQBO and SC-min/wQBO), the LDA results in a single pattern and an associated indicial time series. With four predefined groups (e.g., SC-min/wQBO, SC-min/eQBO, SC-max/wQBO, and SC-max/eQBO), there are three spatial patterns, $P_1(\mathbf{x})$, $P_2(\mathbf{x})$, and $P_3(\mathbf{x})$, which distinguish the observations of each of the groups from those in the other groups. Associated with each spatial patterns is its time series index, $C_1(t)$, $C_2(t)$, and $C_3(t)$. The first pattern, $P_1(\mathbf{x})$, yields the largest separability measure, $\mathcal{R}_1 = \max\mathcal{R}$ as described above. The second pattern, $P_2(\mathbf{x})$, is the pattern that maximizes $\mathcal{R}$ subject to the constraint that its associated index, $C_2(t)$ is uncorrelated with the previous index, $C_1(t)$. This yields a separability measure $\mathcal{R}_2 < \mathcal{R}_1$. Note that $P_2(\mathbf{x})$ is not required to be orthogonal to $P_1(\mathbf{x})$. Similarly, $P_3(\mathbf{x})$ is the pattern maximizing $\mathcal{R}$ such that $C_3(t)$ is uncorrelated with both $C_1(t)$ and $C_2(t)$, yielding a separability measure $\mathcal{R}_3 < \mathcal{R}_2 < \mathcal{R}_1$. A more detailed discussion of the technique can be found in the appendixes of this paper and in Schneider and Held (2001) and Ripley (1996).

It is important to note that the apparent separation, as measured by $\mathcal{R}$, is not a robust feature of the analysis. It is biased to large values and, in particular, is dependent upon the choice of truncation parameter used to reduce the degrees of freedom in the dataset prior to performing the LDA (see below and appendix B for more details on this parameter). As such, it should not be used directly to test the statistical significance of our results. Therefore, to test the statistical robustness of each analysis, we perform a bootstrap Monte Carlo test using 10 000 synthetic datasets constructed by resampling with replacement our set of observations while preserving the group structure and truncation parameter of the original analysis. In other words, we randomly choose observations from the original dataset and assign them to a group regardless of their original physical state while preserving the number of observations in each group. For an example, we might need nine randomly chosen years to put into the SC-min/eQBO category. One year randomly picked could be 1990, which originally was an SC-max/eQBO year. "With replacement" means that the next time we pick a year, all years, including 1990, are again available to be picked. This assures that each random choice is drawn from the same distribution. An LDA is then performed on each synthetic dataset using the same truncation parameter as that used in the original (observed) LDA to create a distribution of variance ratios. The percentile of the observed variance ratio within the distribution of synthesized variance ratios denotes the statistical significance of that observed ratio.

Since the time series for each variable (grid point) exhibit a high degree of colinearity, it is desirable to perform some sort of spatial smoothing to reduce the number of spatial degrees of freedom (dof) prior to performing the LDA. In fact, for short data records such as those we study here, the number of observa-

tions (times) is smaller than the number of variables (temperatures at different latitudes); in this case, a LDA based on temporal groups is ill-posed and a regularization of the dataset is required. In this study, we use a truncated empirical orthogonal function (EOF) expansion of the dataset to perform this task (i.e., we reconstruct a smoothed dataset using only the leading-r EOFs). The technical details on how to choose the truncation level, r, are discussed in appendix B. Briefly, there exists for most analyses a range for r within which the results are qualitatively similar. Since the EOFs are sorted by percentage of variance captured, if r is chosen too small, the EOFs that contain variance associated with the desired behavior will be excluded from the truncated dataset and a poor separation (not statistically significant) will be found. This is particularly true for smaller amplitude signals such as the solar cycle response. If r is chosen too large relative to the amount of data, artificially high separation will be seen, caused by being nearly orthogonal to all the within-group variance of the groups; however, this case will also have poor statistical significance under the test described above since this orthogonality can be achieved by random assignations of observations to the groups. Between these extremes lies a range for r for which reasonable and statistically significant separations can be found. The exact choice of r within this range is more subjective.[2] For smaller r (more heavily truncated), the LDA results in a poorer separation but in the capture of a larger percentage of the variance of the original (untruncated) datasets. As r increases, the separation improves but less variance is captured.

3. Datasets

We use data from the National Centers for Environmental Prediction–National Center for Atmospheric Research (NCEP–NCAR) reanalysis from the late winter of 1953/54 to the late winter of 2004/05; the starting year was dictated by the availability of the equatorial wind above Singapore, which we use for the QBO index. (This dataset was kindly provided to us by B. Naujokat for the period 1953–2005.) We consider the mean temperature in the 10–50-hPa layer, as represented by the difference between the 10-hPa and 50-hPa geopotential height surfaces. We use zonally averaged monthly mean data on a 2.5° latitude grid restricted to latitudes north of 30°N (inclusive). The Tropics are avoided because of the predominance of the equatorial QBO pattern there; inclusion of the Tropics could inflate the statistical significance of our results, and we are primarily interested the polar response. The fields are scaled at each grid point by the square root of the cosine of the latitude to account for the change in the area represented by each grid point. The analysis reported here was carried out using the average of the February and March monthly means.[3] A nonuniform trend, obtained by a cubic polynomial fit to each zonally averaged time series, is removed; Hu and Tung (2002) have found that the NCEP temperatures in the polar stratosphere have a cooling trend during the last 30 years only. The resulting annual datasets are centered by subtracting the mean height difference at each grid point from the corresponding time series.

Observations are assigned to one of four groups: SC-max/wQBO, SC-min/wQBO, SC-max/eQBO, and SC-min/eQBO, based on the phases of the indices used in each analysis. For a solar cycle index, we use the monthly means of the observed daily noon 10.7-cm Solar Radio Flux for 1949–2005 as measured by the National Research Council of Canada at Ottawa/Penticton [data available online at NOAA's National Geophysical Data Center (http://www.ngdc.noaa.gov/stp/SOLAR/ftpsolarradio.html)]. The SC-max (-min) months are defined as those with fluxes greater (less) than 140 (125) solar flux units (sfu); the mean flux was 132.5 sfu. Directly measured total solar irradiances are not available for the first half of our data period. We use the Singapore wind at 30 hPa to define the eQBO and wQBO phases. Since the QBO wind changes sign during the winter of some SC–max years (Salby and Callaghan 2004), we use a 4-month average (December–March). As pointed out by Newman et al. (2001) and Hu and Tung (2002), the warming during a particular month in winter is caused by the cumulative effects of wave forcing in all previous months during the same winter. The westerly (easterly) phase is defined by Singapore wind speeds greater (less) than 4.0 (-4.0) m s^{-1}. Table 1 show the assignation of the observations to the four groups defined by the phases of the SC and QBO.

We have found that the correlation of the mean temperature in the 10–50-hPa layer with the solar cycle is different in early winter (centered in November, but can include October and December) from the correlation found in late winter (centered in February, but can include January and March). Therefore, the selection of

[2] With longer data records (more observations), traditional regularization techniques, such as generalized cross-validation, can be employed to choose an appropriate truncation.

[3] Separate analyses (not shown) were performed for individual months and for other 2-month and 3-month averages in late winter. While they all showed qualitatively similar results, the February–March average had the largest statistical significance.

TABLE 1. Grouping of Feb–Mar observations by the phases of the QBO index (30-hPa Singapore wind, Dec–Mar average) and the SC index (10.7-cm flux). Years with QBO index between −4.0 and 4.0 m s^{-1} or solar index between 125 and 140 sfu were excluded.

	SC-min	SC-max
wQBO	1955, 1962, 1964, 1976, 1985, 1986, 1988, 1995, 1997	1958, 1967, 1978, 1981, 1991, 2000, 2002
eQBO	1954, 1963, 1966, 1975, 1977, 1987, 1994, 1996, 1998	1959, 1968, 1970, 1982, 1989, 1990, 2001

months in late winter for this study is important. Early winter results will be presented separately in another paper. Because of the large variability in the Northern Hemisphere polar vortex during winter, multiple-month averaging is needed to achieve statistically confident results. Three-month averaging is usually best, but we have attempted to better pinpoint the timing by using 1-month and 2-month averages. In general, early winter results can be significant if the chosen period does not include much influence from SSWs, which can occur as early as December in some years. Late winter results are similar as long as the averaging period includes February.

Two major volcanoes erupted during our period: El Chichón in March 1982 and Mt. Pinatubo in June 1991. Since temperature data in February–March is used in this study, the years immediately following the eruptions, 1983 and 1992, are excluded to avoid the contamination by the warming caused by volcanic aerosols, in the stratosphere. Results are qualitatively similar if these years are not dropped except that 1983 would appear as an outlier in most analyses.

4. Four-group analysis

We attempt to discriminate the behavior of the Northern Hemisphere stratosphere during these years among these four groups using the method of linear discriminant analysis. For a four-group analysis, there exist three discriminants consisting of the spatial patterns that best distinguish each group from the other groups and of an associated index time series for each pattern. The index time series are denoted $C_1(t)$, $C_2(t)$, and $C_3(t)$, while the associated patterns are denoted $P_1(\mathbf{x})$, $P_2(\mathbf{x})$, and $P_3(\mathbf{x})$, respectively. The patterns, the P, are scaled such that their values at the North Pole are 1. Therefore, the values of the associated time series, the C, represent the temperature at the North Pole in kelvin; the origin of the C is the mean temperature of all data points. This analysis was performed using an $r = 12$ truncation; statistically significant separations

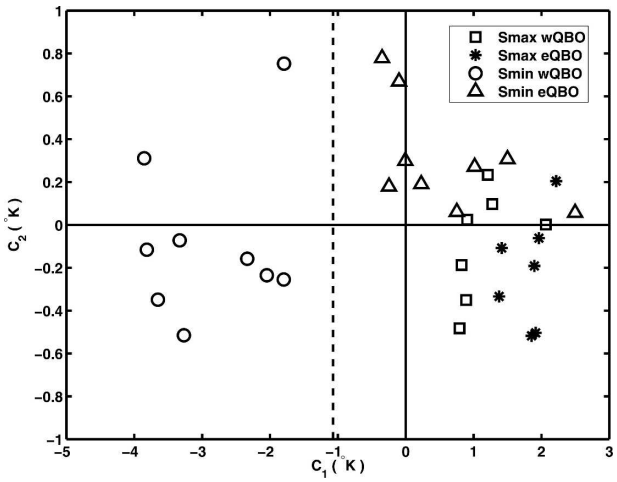

FIG. 4. Scatterplot of the first two canonical variates, C_1 and C_2, from the four-group LDA for the February–March average of the zonally averaged difference between 10-hPa and 50-hPa geopotential surfaces from 1954 to 2005. Grouping based on both the SC and 30-hPa equatorial QBO indices.

were achieved for $7 \leq r \leq 16$ with qualitatively similar results. Figure 4 is a scatterplot of the projection of the years onto the first two discriminant patterns [i.e., $C_2(t)$ versus $C_1(t)$]. It is seen that $C_1(t)$, shown as the horizontal axis, clearly isolates the SC-min/wQBO state from the other three states; the separability measure for $C_1(t)$, $\mathcal{R}_1$, is approximately 8.4. The vertical axis, $C_2(t)$, appears to separate the pure QBO perturbation from the SC perturbations, but the difference in polar temperature is rather small [note the decimal scale for $C_2(t)$]. The second and third discriminants (not shown) distinguish among the other three states, with $\mathcal{R}_2 \approx 0.6$, $\mathcal{R}_3 \approx 0.2$. The variances in the polar regions associated with $P_2(\mathbf{x})$ and $P_3(\mathbf{x})$ are much smaller than the polar variance associated with $P_1(\mathbf{x})$; the second and third discriminants contribute mostly mid- and low-latitudinal variance. The rather small polar variances associated with $P_2(\mathbf{x})$ and $P_3(\mathbf{x})$, physically imply that the perturbations of the polar vortex associated with eQBO and SC-max may be quite similar in spatial pattern. The magnitude of the perturbation (warming) from the least-perturbed state by the eQBO and SC-max appears to be comparable. It is perhaps important to point out that the separate perturbations by the eQBO and SC-max are not additive. That is, the SC-max/eQBO perturbation is not twice that of either the SC-max/wQBO or SC-min/eQBO perturbations. In fact they are all of roughly comparable magnitudes. In particular, the SC-max and eQBO perturbations do not cancel each other out when combined. Figure 5a shows the spatial pattern of the first discriminant, $P_1(\mathbf{x})$, with a warming of the polar vortex when $C_1(t)$ is positive.

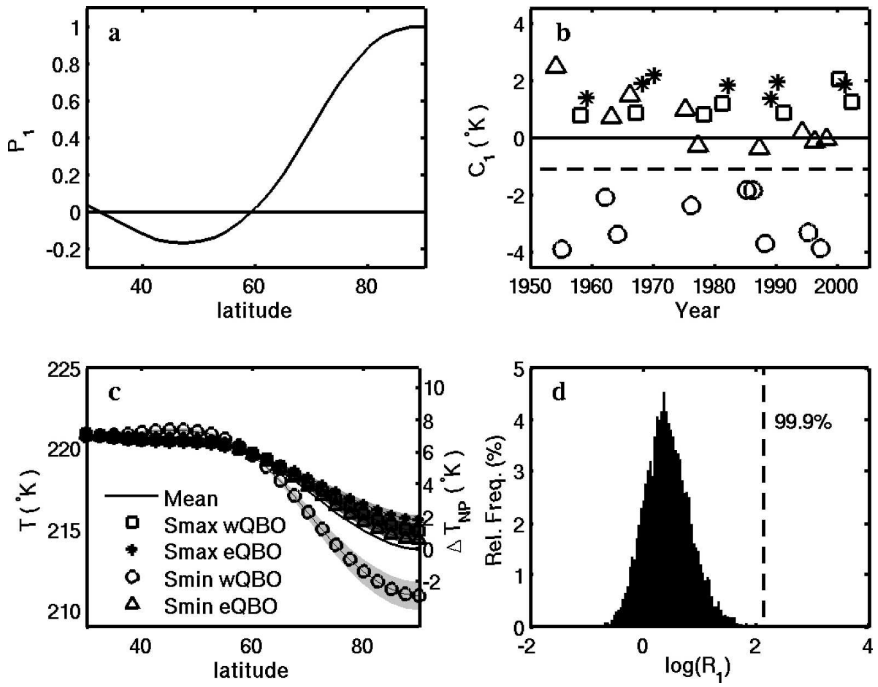

FIG. 5. Results of a four-group LDA for the February–March average of the zonally averaged difference between 10-hPa and 50-hPa geopotential surfaces from 1954 to 2005. Grouping based on both the SC and 30-hPa equatorial QBO indices: (a) first discriminant pattern, $P_1(\mathbf{x})$, (b) first time series index: $C_1(t)$. (c) Mean state (unmarked line) plus group-mean projections onto $P_1(\mathbf{x})$ for all four groups. Shaded regions show 1σ projections for all groups, and (d) Monte Carlo distribution of variance ratios showing percentile of observed variance ratio, $\mathcal{R}_1$, for the first discriminant.

Figure 5b shows $C_1(t)$ as a function of year and shows that values of $C_1 \lesssim -1$ K nicely distinguish the SC-min/wQBO years as being colder and less perturbed when compared to the other three groups. It does not appear to distinguish SC–max perturbations from eQBO perturbations, but shows that these two perturbations are comparable in magnitude. Comparing the mean of $C_1(t)$ during the unperturbed SC-min/wQBO years to the mean during perturbed years of the other three groups (i.e., SC-min/eQBO, SC-max/wQBO, and SC-max/eQBO) we can see that there is zonal-mean warming of the pole of approximately 4.2 K. This can be more readily seen in Fig. 5c, which gives an alternate presentation of the information already contained in Figs. 5a,b. It shows the mean anomaly associated with $P_1(\mathbf{x})$ for each group superimposed on the climatology (overall mean). Mean anomalies were determined by multiplying $P_1(\mathbf{x})$ by the mean of C_1 within each group of years. Shaded regions denote the one standard deviation projections within each group. Figure 5c shows that the perturbed years, when projected onto this coherent spatial pattern, are well separated from the SC-min/wQBO years. A Monte Carlo test, Fig. 5d, of the four-group variance ratio shows that the observed separation measure, $\mathcal{R}_1$, is significant at a 99.9% confidence level. This is the first key result of our present work:

> There exists a least-perturbed state of the polar vortex during SC-min/wQBO that is statistically significant (at the 99.9% confidence level); the SC-max and eQBO each provides a comparable warming perturbation from this state. The combined perturbation from the SC-max and eQBO also warms the pole by a similar magnitude.

It should be noted that these figures do not capture the full variability of the polar vortex, only that portion which projects onto $P_1(\mathbf{x})$. Furthermore, $P_1(\mathbf{x})$ does not distinguish between perturbations of the polar vortex during the SC-max from those during the easterly QBO. We are also interested in finding out the spatial structure of the combined perturbation. Therefore, we will perform a series of two-group LDAs to isolate SC effects from QBO effects. Unlike previous studies, we obtain statistically significant results when we define perturbations as relative to the least-perturbed state.

5. Solar cycle perturbation

A pure solar maximum perturbation relative to the least-perturbed state is obtained when we perform a

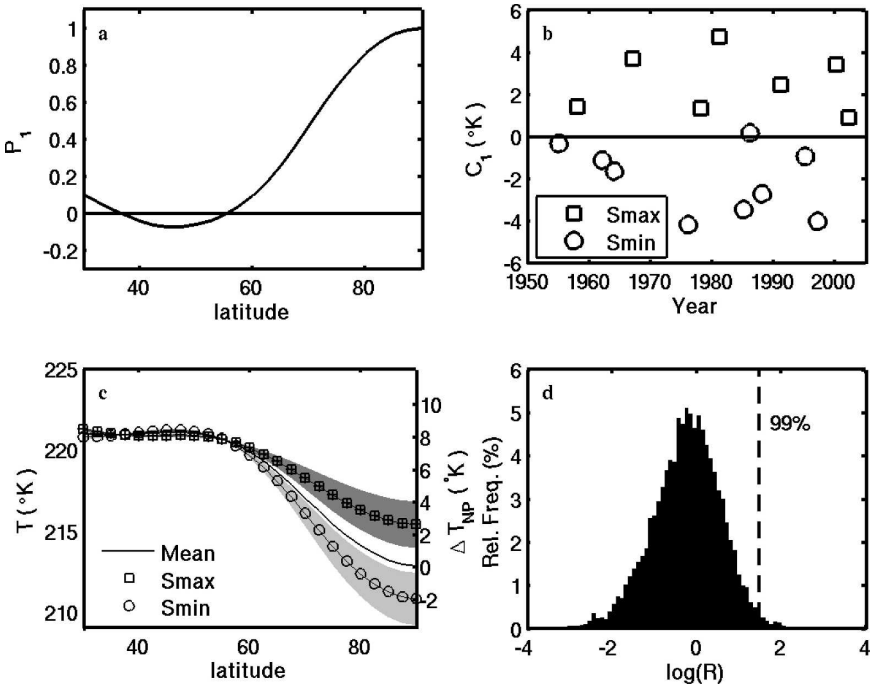

FIG. 6. As in Fig. 5 except that the LDA is based on the phase of the solar cycle during westerly QBO years only: the dark □ denotes the SC-max/wQBO group and the light ○ denotes the SC-min/wQBO group.

two-group LDA discriminating the SC-max/wQBO years from the SC-min/wQBO years. We use the same two-month temporal average (February–March) in geopotential height and the same QBO and SC indices as used in the previous four-group analysis. Figure 6 shows the results using an $r = 5$ truncation; similar analyses with $4 \leq r \leq 7$ all have statistically significant separations with confidence levels above 95%.[4] The spatial pattern shown in Fig. 6a is similar to that obtained in the four-group analysis (see Fig. 5a) with a warming of the pole and a cooling of the midlatitudes during the SC-max. It is previously known that the SC-max warms the stratosphere radiatively over all latitudes (Coughlin and Tung 2004) and that the ozone heating is strongest over the equatorial stratosphere (Shindell et al. 1999; Haigh 1999; Hood et al. 2001). The dynamical part of the pattern should be interpreted as that obtainable by offsetting Fig. 6a by 0.5–1.0 K. The resulting pattern is then seen as a larger polar warming balanced by a smaller cooling over a wider region in the mid and low latitudes. This is then consistent with the warming and cooling structure resulting from the deposition of easterly momentum in the high-latitude stratosphere (Garcia 1987) and, in particular, with the predicted pattern during a SSW (Matsuno 1971).

The mean warming of the pole is approximately 4.6 K (see Fig. 6c) as measured by the difference between the mean of the SC-max group and the mean of the SC-min group. A more common measure is the cycle peak-to-peak difference, which is almost 7.2 K.[5] The Monte Carlo test shows that the SC-max years are separated from the SC-min years by this spatial pattern at the 99% confidence level.

The warming found here is substantially smaller than the 14-K warming reported previously by LvL88 and Salby and Callaghan (2000). However, these previous results are for a single point, the North Pole, and at the level of the maximum, 10 hPa, while ours is the polar value of a coherent mean meridional structure, averaged over a thick layer, 10–50 hPa. This is the second key result of this study:

> It establishes the statistical significance (at the 99% confidence level) of a pattern of warming due to the solar cycle influence on the stratospheric temperature. Its spatial structure is consistent with that of SSW.

[4] Since the two-group LDAs use roughly half as many observations as the four-group LDA, we expect the dof and therefore the appropriate truncation, r, to be similarly reduced.

[5] The peak-to-peak difference is obtained by dividing the mean difference by 0.636, the mean of $|\sin(t)|$.

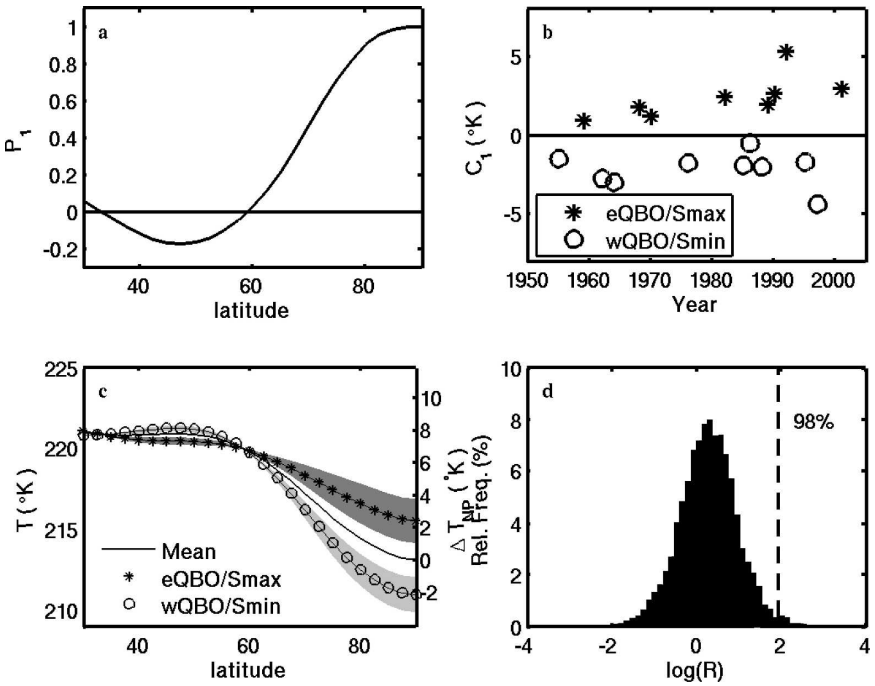

FIG. 7. As in Fig. 5 except that the LDA is based on the combined SC-max/eQBO perturbation: the dark * denotes the SC-max/eQBO group and the light ○ denotes the SC-min/wQBO group.

6. Combined solar cycle and QBO perturbation

Figure 7 show the result of a two-group LDA for the combined perturbation from the least-perturbed state (i.e., SC-max/eQBO versus SC-min/wQBO). Remarkably, the separation is highly significant at the 98% confidence level; a $r = 7$ truncation was used. (The analysis with $r = 8$ is also statistically significant at above the 95% level.) The spatial pattern and magnitude of the polar warming by the combined perturbation of the SC-max and eQBO is about the same as that of the SC-max alone.[6] This suggests that, if the warming is caused by the induced occurrence of SSW, once a SSW occurs by the effect of the eQBO (or SC-max), the additional perturbation by the SC-max (or eQBO) does not further warm the polar stratosphere by inducing another SSW during the same late winter. Thus, the effects of the eQBO and SC-max perturbations are not additive. The third key results of this study is:

> The combined perturbation from the solar cycle and from the QBO relative to the least-perturbed state is a polar warming with a latitudinal structure consistent with a SSW. There is no cancellation of the SC effect by the QBO effect, or vice versa. This result is statistically significant at the 98% confidence level.

7. QBO perturbation

Figure 8 shows the result of a two-group LDA discriminating SC-min/eQBO years from SC-min/wQBO years, using an $r = 7$ truncation. The spatial pattern of polar warming and smaller mid- and low-latitude cooling is consistent with the structure of a SSW. The magnitude of the polar warming is similar to the previous analyses, with a 3.8-K warming from mean-wQBO to mean-eQBO years and approximately 6-K warming for peak to peak. However, the result is just barely significant at the 90% confidence level; no other truncations result in a statistically significant separation. QBO results are sensitive to the choice of vertical level(s) in defining its phase; this is discussed further below.

8. Conclusions

Our results can be summarized in a four-quadrant diagram (see Fig. 9). The arrows indicate the direction of polar warming.

[6] It is thus not surprising that traditional analyses attempting to find the difference between eQBO and wQBO during the SC–max yields confusing results.

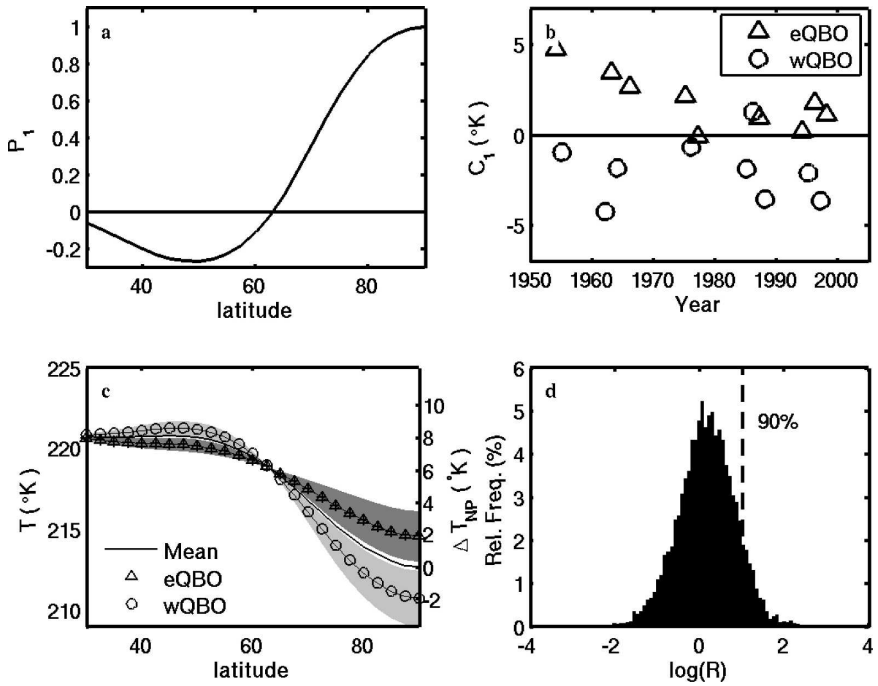

FIG. 8. As in Fig. 5 except that the LDA is based on the phase of the QBO during solar minimum years only: the dark Δ denotes the SC-min/eQBO group and the light ○ denotes the SC-min/wQBO group.

1) There is a very well separated state of the polar stratosphere, which occurs during the confluence of a solar cycle minimum and westerly QBO; we denote this state as the "least-perturbed state." This state is statistically distinguished from all other (perturbed) states at more than the 99% confidence level.
2) Relative to this least-perturbed state, a solar maximum warms the pole by a mean of approximately 4.6 K (7.2 K peak to peak) during the wQBO and also during the eQBO; both results are statistically significant at more than the 95% confidence level. The statistically significant discriminant spatial patterns of the warming take the form of the structure associated with sudden stratospheric warmings.
3) The above result shows that the solar cycle warms the polar stratosphere by the same amount *regardless* of the phase of the QBO. It then follows that the stratification of the data according to the phase of the QBO is not necessary in order to see the SC response, except for its necessity in defining the reference state.
4) Relative to the least-perturbed state, easterly QBO also warms the pole by approximately 4 K but at a lower confidence level of 90%. This lowered significance may be due to the fact that the preconditioning of the SSW by a QBO may depend on the equatorial wind at several height layers (Gray et al. 2001; Gray 2003). The spatial structure is also similar to that of the SSW.
5) Statistically significant results are obtained when perturbations are measured relative to the least-perturbed state. Previous confusion over possible reversals of solar cycle warming or of the Holton–Tan QBO mechanism are now seen to arise from the comparison of one perturbed state with another perturbed state. These differences are not statistically significant due to the similarity of each of the perturbations. In Fig. 9, the differences between perturbed states are denoted by dashed arrows. Two-group LDAs between the perturbed states (not shown) show that there is an approximately 0.4-K warming from SC-min/eQBO to SC-max/eQBO (i.e., SC perturbation during eQBO) and an approximately 0.3-K cooling from SC-max/wQBO to SC-max/eQBO (i.e., QBO perturbation during the SC-max). The latter result (the small cooling) has been interpreted as a reversal of the Holton–Tan mechanism; however, both of these results have very low statistical significances and so the signs of the warming should not be taken seriously.

We now discuss some previous work related to understanding the physical mechanisms of the cause of the

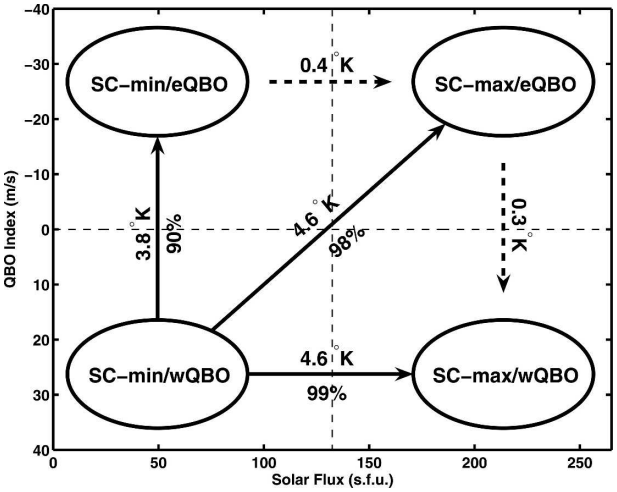

FIG. 9. Summary schematic showing that the state of the SC-min/wQBO is the least-perturbed state. Solid arrows indicate the mean warming of the pole for the perturbed states relative to this state. Confidence levels for the associated LDAs are also shown. Dashed arrows indicate the small difference found between perturbed states, which are not statistically significant. All results are for the February–March average of the 10–50-hPa layer mean temperature.

solar cycle warming. The polar stratosphere in the Northern Hemisphere has large interannual variability during winter. The lower stratosphere is much warmer than can be explained from radiative considerations alone, especially during the polar night. The difference in temperature between the observed polar temperature and its radiative equilibrium value is now understood to be provided by the dynamical heating associated with the planetary waves forced near the surface and propagated upward into the lower stratosphere, where they sometimes break and as a consequence deposit their momentum and heat energy. The most spectacular of these breaking events occur in sudden stratospheric warmings, when, in the course of one week, the temperature of the polar stratosphere can increase by a few tens of kelvins. It is well known that the easterly phase of the QBO preconditions the polar stratosphere for midwinter SSWs (McIntyre 1982; Butchart et al. 1982; Smith 1989; Tung and Lindzen 1979a,b; Tung 1979). Dunkerton et al. (1988) found that none of the observed SSWs (up to 1987) had occurred while the QBO was in a deep westerly phase. While the polar vortex during westerly years of the QBO tends to be less perturbed and generally less prone to SSW occurrences, there are nevertheless exceptional years when SSWs do occur (Labitzke 1982; NH97); these occur during the solar cycle maximum. In a recent update, Labitzke et al. 2006) found that out of 11 cases of midwinter SSWs during the wQBO, none occurred during the SC minimum and 10 occurred during the SC maximum. The modeling work of Gray et al. (2004), Matthes et al. (2004), Pascoe et al. (2006), and Palmer and Gray (2005) offers some support to this observed result, possibly as caused by changes in the timing of SSWs.

We observe warming of the polar stratospheric vortex during solar cycle maxima as compared to SC minima during late winter. This SC-induced warming during the westerly QBO is large, about 7 K in the zonal mean temperature over the broad layer of 10–50 hPa and appears to be due to dynamical heating arising from SSWs. Although individual SSWs warm the pole by more than 7 K, their typical duration is approximately one week and therefore a smaller average warming is expected from the multimonth averages used here. The energy for SSWs originates from the denser troposphere; therefore, a modulation of the frequency of SSW occurrence can provide a powerful dynamical amplifier for the radiative effect of the solar cycle. Thus, we find, as pointed out by Labitzke (1982) and NH97, that the exceptional years when SSWs occur in the westerly phase of the QBO are years of a solar cycle maximum. Our result puts this on firmer statistical ground. We have isolated the rather large amplitude of the SC effect on the polar temperature during February–March. Since this time is during polar night, it is unlikely to be a radiative effect. A new result of this study is that the SC-max/eQBO perturbation is comparable to the SC–max perturbation alone and is statistically significant. There is no "reversal" of the solar effect by the QBO or vice versa. It appears that each perturbation enhances the frequency of the SSWs, but that the combined perturbation does not double it during the same period.

Balachandran and Rind (1995) and Rind and Balachandran (1995) performed a general circulation model calculation with imposed QBO and an exaggerated UV variation of 10% for wavelengths less than 0.3 μm. During the wQBO phase, the solar cycle UV effect produced a warming of the polar region of 8 K near 30 km. This model result is consistent with our observational finding presented here. They attributed the difference to the change to the planetary wave propagation caused by the effect of the UV heating on the vertical shear of the zonal wind. They also found that during the SC-min, the eQBO produces a polar warming maximum of slightly less than 8 K located lower down, at 20 km. In the region of 20–30 km analyzed by us, the warming reported by them, is smaller, about 3–4 K. They attributed this warming to the effect of eQBO on the meridional gradient of the zonal wind; the resulting change in the index of refraction converges

planetary wave energy toward the lower-stratosphere high latitudes.

In the tropical stratosphere, which is not the focus of this paper, the interaction of the SC and QBO may be more complicated because of the possible dual radiative–dynamical role played by ozone in this region of maximum exposure to the sun's radiation. (Ozone's absorption of solar ultraviolet radiation may manifest SC variability as a direct radiative heating. Also, by driving a diabatic meridional circulation, such heating may affect the descent rate of the equatorial QBO.) Salby and Callaghan (2000) proposed that the length of the westerly phase of the equatorial QBO may be affected by the solar cycle, being longer during solar minimum conditions. Soukharev and Hood (2001) found the lengths of both QBO phases to be longer during the solar minimum, although the data record is too short to permit statistical tests. Gray et al. (2001) and Gray (2003) suggested that, while the occurrence of SSW is sensitive to the direction of the equatorial wind in the lower stratosphere in early winter, in mid and late winter it is more sensitive to the wind direction in the equatorial upper stratosphere (above 40 km). Since the solar cycle effect is stronger in the upper stratosphere, this may be the route that the SC effect is introduced to the polar stratosphere. This may also explain why the QBO warming is less robust in late winter. Kodera and Kuroda (2002) suggested a conceptual model, based on their analysis of 20 years of data, that the radiatively controlled state at the tropical stratopause level lasts longer during solar maximum, and the stratopause subtropical jet reaches a higher speed. Then, through interaction with the propagation of planetary waves, such a radiative effect is transmitted downward and poleward in the course of months into the polar lower stratosphere as a modulation of an internal mode of variation of the zonal mean winds in the winter stratosphere. Since our emphasis is on statistical tests, which requires reasonable quality data of more than four solar cycles, which are not available above 10 hPa, these proposals concerning the upper stratosphere are not pursued in the present work. The issue concerning the timing for the occurrence of the SSW raised by Gray et al. (2004) will be further investigated in a separate paper contrasting early winter with late winter behavior. The current paper concerns only the February–March average. There were also reports that the power spectrum of the solar flux contains a small secondary peak near the QBO period (Shapiro and Ward 1962; Soukharev and Hood 2001). Although this "QBO" peak in the solar spectrum may be statistically significant, its magnitude is two orders of magnitude smaller than the main peak at 11 yr, so it is unlikely that this is the cause of the SC–QBO interaction seen in the atmospheric data, at least not in the lower stratosphere.

Acknowledgments. We thank Tapio Schneider for his contribution in useful discussions regarding discriminant analysis and statistical testing. We also thank Katie Coughlin for discussions on the QBO–SC interactions. We are grateful to three anonymous reviewers for their constructive comments. The research was supported by the National Science Foundation under Grant ATM-3 32364.

APPENDIX A

Discriminant Analysis

For a set of multivariate observations, $\mathbf{X}$, linear discriminant analysis seeks a linear combination, $\mathbf{Xu}$, for which the separation between predefined groups is maximized. The separation can be defined as the ratio of between-group variance to either within-group or total variance. Suppose that we have a centered dataset, $\mathbf{X}(n \times p)$, consisting of n observations of p variables, $\mathbf{x}_i$ $(1 \times p)$. Centering is performed by removing the mean observations from the original data. Furthermore, suppose that we can partition the observations into g groups, G_j $(j = 1, \ldots, g)$, with n_j observations in each group. Let $\boldsymbol{\mu}_j$ $(1 \times p)$ denote the mean of the observations in G_j and $[i]$ denote the group of observation $\mathbf{x}_i$. We can define the group matrix $\mathbf{G}(n \times g) = [g_{ij}]$, where $g_{ij} = 1$ if $j = [i]$ and equals 0, otherwise. Note that $\mathbf{G}^{\mathrm{T}}\mathbf{G} = \mathrm{diag}(n_j)$. Let

$$\mathbf{M}(g \times p) = (\mathbf{G}^{\mathrm{T}}\mathbf{G})^{-1}\mathbf{G}^{\mathrm{T}}\mathbf{X} = [\boldsymbol{\mu}^{\mathrm{T}}]$$

be the matrix of group means. Then the within-group covariance matrix can be defined as

$$\boldsymbol{\Sigma}_w(p \times p) = \frac{1}{n - g}(\mathbf{X} - \mathbf{GM})^{\mathrm{T}}(\mathbf{X} - \mathbf{GM}),$$

the between-group-covariance matrix as

$$\boldsymbol{\Sigma}_a(p \times p) = \frac{g}{n(g - 1)}(\mathbf{GM})^{\mathrm{T}}(\mathbf{GM})$$

$$= \frac{g}{n(g - 1)}\mathbf{X}^{\mathrm{T}}\mathbf{G}(\mathbf{G}^{\mathrm{T}}\mathbf{G})^{-1}\mathbf{G}^{\mathrm{T}}\mathbf{X},$$

and the total-covariance matrix as

$$\boldsymbol{\Sigma}_t(p \times p) = \frac{1}{n - 1}\left((n - g)\boldsymbol{\Sigma}_w + \frac{g}{n(g - 1)}\boldsymbol{\Sigma}_a\right).$$

A separability measure for any linear combination $\mathbf{Xu}$ can then be defined by

$$\gamma = (\mathbf{u}^{\mathrm{T}}\boldsymbol{\Sigma}_t\mathbf{u})^{-1}(\mathbf{u}^{\mathrm{T}}\boldsymbol{\Sigma}_a\mathbf{u}).$$

The canonical variates, $\mathbf{c}_k = \mathbf{X}\mathbf{u}_k$ ($k = 1, \ldots, g-1$), are the linear combinations which optimize γ, which implies that $\mathbf{u}_k$ must satisfy

$$\partial\gamma/\partial\mathbf{u} = 2(\mathbf{u}^T\boldsymbol{\Sigma}_t\mathbf{u})^{-1}\boldsymbol{\Sigma}_a\mathbf{u} - 2(\mathbf{u}^T\boldsymbol{\Sigma}_t\mathbf{u})^{-2}(\mathbf{u}^T\boldsymbol{\Sigma}_a\mathbf{u})\boldsymbol{\Sigma}_t\mathbf{u} = 0.$$

Therefore, the $\mathbf{u}_k$ must satisfy $\boldsymbol{\Sigma}_t^{-1}\boldsymbol{\Sigma}_a\mathbf{u}_k = \gamma_k\mathbf{u}_k$, where the γ_k are the ordered eigenvalues of $\boldsymbol{\Sigma}_t^{-1}\boldsymbol{\Sigma}_a$ and the $\mathbf{u}_k$ are the associated right eigenvectors. This is equivalent to maximizing $\mathcal{R} = (\mathbf{u}^T\boldsymbol{\Sigma}_w\mathbf{u})^{-1}(\mathbf{u}^T\boldsymbol{\Sigma}_a\mathbf{u})$. The canonical variates, $\mathbf{c}_k$, can be arbitrarily normalized [e.g., $\text{var}(\mathbf{c}_k) = \mathbf{u}_k^T\boldsymbol{\Sigma}_t\mathbf{u}_k = 1$]. The discriminating patterns are then defined as the regression coefficients, $\mathbf{p}_k = (\mathbf{u}_k^T\boldsymbol{\Sigma}_t\mathbf{u}_k)^{-1}\boldsymbol{\Sigma}_t\mathbf{u}_k$ of the centered data, $\mathbf{X}$, onto the $\mathbf{c}_k$.

APPENDIX B

Truncation

Since the number of variables (grid points) exceeds the number of observations (times), the system to be analyzed is overdetermined and the LDA is ill posed. Therefore, the analysis must be regularized (Schneider and Held 2001; Hansen 1998). The LDA is performed using a truncated principal component representation of the dataset; keeping only the leading r modes. Figure B1a depicts the dependence of the first variance ratio, $\mathcal{R}_1 = \upsilon_b/\upsilon_w$, on the choice of the truncation parameter, r. This example is drawn from the four-group LDA performed in section 4. The between-group variance, υ_b, and within-group variance, υ_w, are also independently shown. We can see that as r increases, both $\mathcal{R}_1$ and υ_b increase while υ_w decreases. As the number of degrees of freedom grow, the discriminant is able to capture more of the large-scale variance between the groups while simultaneously becoming more orthogonal to the within-group variability. However, relative importance of υ_b and υ_w to $\mathcal{R}_1$ does not stay consistent. Initially as r grows, $r \leq 12$, the change in $\mathcal{R}_1$ is caused by changes in both υ_b and υ_w. For $r \geq 13$, increases in $\mathcal{R}_1$ are caused primarily by decreases in υ_w, which may not be desirable. This can seen more clearly in Fig. B1b, which depicts the ratio of $\mathcal{R}_1$ for neighboring values of r: $\Delta\mathcal{R}_1(r) = \mathcal{R}_1(r)/\mathcal{R}_1(r-1)$. Similar ratios for υ_b and υ_w are also shown.

If r is too small, the model is missing variability important to the discrimination of the groups. This can be seen in rapid growth of $\mathcal{R}_1$ for $r \leq 7$. On the other hand, if r is too large, the model is overfitting the data. Since the number of observations in each group is small, choosing a large r gives the model enough degrees of freedom to find a pattern nearly orthogonal to the within-group variability. This rapid decrease in υ_w and associated rapid increase in $\mathcal{R}_1$ is most likely an artifact of the small sample size. For this example, choosing $7 \leq r \leq 13$ seems appropriate; we depicted $r = 12$ in this study.

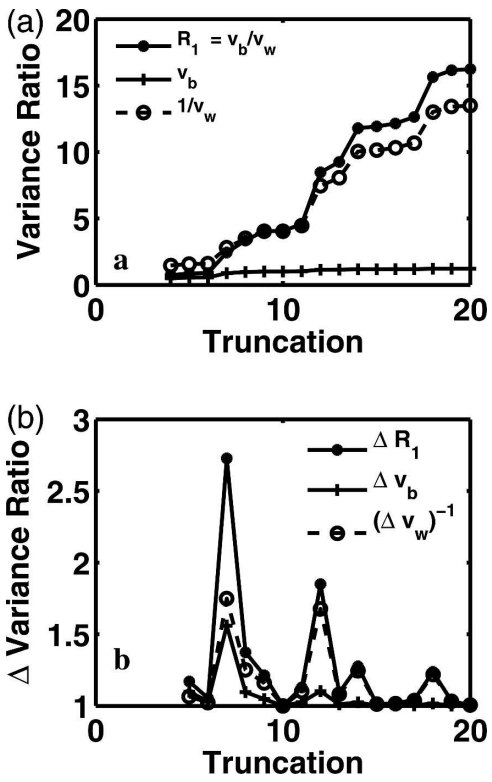

FIG. B1. (a) Dependence of the first variance ratio, $\mathcal{R}_1$, on the EOF truncation number, r, for the four-group LDA from section 4. Between-group variance, υ_b, and the inverse of the within-group variance, υ_w^{-1}, also shown. (b) Ratio of neighboring values for the variances shown [e.g., $\Delta\mathcal{R}_1 = \mathcal{R}_1(r)/\mathcal{R}_1(r-1)$].

It is important to note that the qualitative nature of results is robust for all but the smallest values of r. It is only the quantitative results, such as the variance ratio and its statistical significance, which depend upon this choice.

REFERENCES

Balachandran, N. K., and D. Rind, 1995: Modeling the effects of UV variability and the QBO on the troposphere–stratosphere system. Part I: The middle atmosphere. *J. Climate,* **8,** 2058–2079.

Butchart, N., S. A. Clough, T. N. Palmer, and P. J. Trevelyan, 1982: Simulations of an observed stratospheric warming with quasigeostrophic refractive-index as a model diagnostic. *Quart. J. Roy. Meteor. Soc.,* **108,** 475–502.

Camp, C. D., and K. K. Tung, 2007: Stratospheric polar warming by ENSO: A statistical study. *Geophys. Res. Lett.,* **34,** L04809, doi:10.1029/2006GL028521.

Coughlin, K. T., and K. K. Tung, 2004: 11-year solar cycle in the stratosphere extracted by the empirical mode decomposition method. *Adv. Space Res.,* **34,** 323–329.

Crooks, S. A., and L. J. Gray, 2005: Characterization of the 11-year solar signal using a multiple regression analysis of the ERA-40 dataset. *J. Climate,* **18,** 996–1015.

Dunkerton, T. J., D. P. Delisi, and M. P. Baldwin, 1988: Distribution of major stratospheric warmings in relation to the quasi-biennial oscillation. *Geophys. Res. Lett.,* **15,** 136–139.

Garcia, R. R., 1987: On the mean meridional circulation of the middle atmosphere. *J. Atmos. Sci.,* **44,** 3599–3609.

Gray, L. J., 2003: The influence of the equatorial upper stratosphere on stratospheric sudden warmings. *Geophys. Res. Lett.,* **30,** 1166, doi:10.1029/2002GL016430.

——, S. J. Phipps, T. J. Dunkerton, M. P. Baldwin, E. F. Drysdale, and M. R. Allen, 2001: A data study of the influence of the equatorial upper stratosphere on northern-hemisphere stratospheric sudden warmings. *Quart. J. Roy. Meteor. Soc.,* **127,** 1985–2003.

——, S. Crooks, C. Pascoe, S. Sparrow, and M. Palmer, 2004: Solar and QBO influences on the timing of stratospheric sudden warmings. *J. Atmos. Sci.,* **61,** 2777–2796.

Haigh, J. D., 1996: The impact of solar variability on climate. *Science,* **272,** 981–984.

——, 1999: A GCM study of climate change in response to the 11-year solar cycle. *Quart. J. Roy. Meteor. Soc.,* **125,** 871–892.

——, 2003: The effects of solar variability on the earth's climate. *Philos. Trans. Roy. Soc. London,* **361,** 95–111.

Hansen, P. C., 1998: *Rank-Deficient and Discrete III-Posed Problems: Numerical Aspects of Linear Inversion. SIAM Monogr.,* Mathematical Modeling and Computation, No. 4, Society for Industrial and Applied Mathematics, 247 pp.

Holton, J. R., and H. C. Tan, 1980: The influence of the equatorial quasi-biennial oscillation on the global circulation at 50 mb. *J. Atmos. Sci.,* **37,** 2200–2208.

——, and ——, 1982: The quasi-biennial oscillation in the northern hemisphere lower stratosphere. *J. Meteor. Soc. Japan,* **60,** 140–148.

Hood, L. L., and B. E. Soukharev, 2003: Quasi-decadal variability of the tropical lower stratosphere: The role of extratropical wave forcing. *J. Atmos. Sci.,* **60,** 2389–2403.

——, ——, M. Fromm, and J. P. McCormack, 2001: Origin of extreme ozone minima at middle to high northern latitudes. *J. Geophys. Res.,* **106,** 20 925–20 940.

Hu, Y. Y., and K. K. Tung, 2002: Interannual and decadal variations of planetary wave activity, stratospheric cooling, and Northern Hemisphere annular mode. *J. Climate,* **15,** 1659–1673.

Kodera, K., 1993: Quasi-decadal modulation of the influence of the equatorial quasi-biennial oscillation on the north polar stratospheric temperatures. *J. Geophys. Res.,* **98,** 7245–7250.

——, and Y. Kuroda, 2002: Dynamical response to the solar cycle. *J. Geophys. Res.,* **107,** 4749, doi:10.1029/2002JD002224.

Labitzke, K., 1982: On the interannual variability of the middle stratosphere during the northern winters. *J. Meteor. Soc. Japan,* **60,** 124–139.

——, 1987: Sunspots, the QBO, and the stratospheric temperature in the north polar-region. *Geophys. Res. Lett.,* **14,** 535–537.

——, 2001: The global signal of the 11-year sunspot cycle in the stratosphere: Differences between solar maxima and minima. *Meteor. Z.,* **10,** 83–90.

——, 2003: The global signal of the 11-year sunspot cycle in the atmosphere: When do we need the QBO? *Meteor. Z.,* **12,** 209–216.

——, 2004: On the signal of the 11-year sunspot cycle in the stratosphere and its modulation by the quasi-biennial oscillation. *J. Atmos. Sol. Terr. Phys.,* **66,** 1151–1157.

——, 2005: On the solar cycle-QBO relationship: A summary. *J. Atmos. Sol. Terr. Phys.,* **67,** 45–54.

——, and H. van Loon, 1988: Associations between the 11-year solar-cycle, the QBO and the atmosphere. Part I: The troposphere and stratosphere in the northern hemisphere in winter. *J. Atmos. Terr. Phys.,* **50,** 197–206.

——, M. Kunze, and S. Brönnimann, 2006: Sunspots, the QBO, and the stratosphere in the north polar region–20 years later. *Meteor. Z.,* **15,** 355–363.

Lean, J., 1991: Variations in the sun's radiative output. *Rev. Geophys.,* **29,** 505–535.

Matsuno, T., 1971: Dynamical model of stratospheric sudden warming. *J. Atmos. Sci.,* **28,** 1479–1494.

Matthes, K., U. Langematz, L. L. Gray, K. Kodera, and K. Labitzke, 2004: Improved 11-year solar signal in the Freie Universitat Berlin Climate Middle Atmosphere Model (FUB-CMAM). *J. Geophys. Res.,* **109,** D06101, doi:10.1029/2003JD004012.

McIntyre, M. E., 1982: How well do we understand the dynamics of stratospheric warmings? *J. Meteor. Soc. Japan,* **60,** 37–65.

——, and T. N. Palmer, 1984: The surf zone in the stratosphere. *J. Atmos. Terr. Phys.,* **46,** 825–849.

Naito, Y., and I. Hirota, 1997: Interannual variability of the northern winter stratospheric circulation related to the QBO and the solar cycle. *J. Meteor. Soc. Japan,* **75,** 925–937.

Newman, P. A., E. R. Nash, and J. E. Rosenfield, 2001: What controls the temperature of the Arctic stratosphere during the spring? *J. Geophys. Res.,* **106,** 19 999–20 010.

Palmer, M. A., and L. J. Gray, 2005: Modeling the atmospheric response to solar irradiance changes using a GCM with a realistic QBO. *Geophys. Res. Lett.,* **32,** L24701, doi:10.1029/2005GL023809.

Pascoe, C. L., L. J. Gray, S. A. Crooks, M. N. Juckes, and M. P. Baldwin, 2005: The quasi-biennial oscillation: Analysis using ERA-40 data. *J. Geophys. Res.,* **110,** D08105, doi:10.1029/2004JD004941.

——, ——, and A. A. Scaife, 2006: A GCM study of the influence of equatorial winds on the timing of sudden stratospheric warmings. *Geophys. Res. Lett.,* **33,** L06825, doi:10.1029/2005GL024715.

Rind, D., and N. K. Balachandran, 1995: Modeling the effects of UV variability and the QBO on the troposphere–stratosphere system. Part II: The troposphere. *J. Climate,* **8,** 2080–2095.

Ripley, B. D., 1996: *Pattern Recognition and Neural Networks.* Cambridge University Press, 403 pp.

Salby, M. L., and D. J. Shea, 1991: Correlations between solar-activity and the atmosphere—An unphysical explanation. *J. Geophys. Res.,* **96,** 22 579–22 595.

——, and P. Callaghan, 2000: Connection between the solar cycle and the QBO: The missing link. *J. Climate,* **13,** 2652–2662.

——, and ——, 2004: Evidence of the solar cycle in the general circulation of the stratosphere. *J. Climate,* **17,** 34–46.

——, ——, and D. Shea, 1997: Interdependence of the tropical and extratropical QBO: Relationship to the solar cycle versus a biennial oscillation in the stratosphere. *J. Geophys. Res.,* **102,** 29 789–29 798.

Schneider, T., and I. M. Held, 2001: Discriminants of twentieth-

century changes in earth surface temperatures. *J. Climate,* **14,** 249–254.

Shapiro, R., and F. Ward, 1962: A neglected cycle in sunspot numbers? *J. Atmos. Sci.,* **19,** 506–508.

Shindell, D., D. Rind, N. Balachandran, J. Lean, and P. Lonergan, 1999: Solar cycle variability, ozone, and climate. *Science,* **284,** 305–308.

Smith, A. K., 1989: An investigation of resonant waves in a numerical model of an observed sudden stratospheric warming. *J. Atmos. Sci.,* **46,** 3038–3054.

Soukharev, B. E., and L. L. Hood, 2001: Possible solar modulation of the equatorial quasi-biennial oscillation: Additional statistical evidence. *J. Geophys. Res.,* **106,** 14 855–14 868.

Tung, K. K., 1979: Theory of stationary long waves. Part III: Quasi-normal modes in a singular waveguide. *Mon. Wea. Rev.,* **107,** 751–774.

——, and R. S. Lindzen, 1979a: Theory of stationary long waves. Part I: Simple theory of blocking. *Mon. Wea. Rev.,* **107,** 714–734.

——, and ——, 1979b: Theory of stationary long waves. Part II: Resonant Rossby waves in the presence of realistic vertical shears. *Mon. Wea. Rev.,* **107,** 735–750.

Wilks, D. S., 1995: *Statistical Methods in the Atmospheric Sciences.* International Geophysics Series, Vol. 59 Academic Press, 464 pp.

Willson, R. C., H. S. Hudson, C. Frohlich, and R. W. Brusa, 1986: Long-term downward trend in total solar irradiance. *Science,* **234,** 1114–1117.