JOURNAL OF GEOPHYSICAL RESEARCH, VOL. 113, D05114, doi:10.1029/2007JD009164, 2008

Solar cycle warming at the Earth's surface in NCEP and ERA-40 data: A linear discriminant analysis

Ka Kit Tung[1] and Charles D. Camp[1,2]

Received 12 August 2007; revised 6 October 2007; accepted 7 December 2007; published 13 March 2008.

[1] The total solar irradiance (TSI) has been measured by orbiting satellites since 1978 to vary on an 11-year cycle by about 0.08%. Because of previous controversies on the reality of solar cycle response at the surface, in this work we discuss the robustness of the solar response with respect to analysis methods, data sets and periods used. Furthermore we concentrate on the globally coherent signal. Two reanalysis data sets are used: one is from National Centers for Environmental Prediction and National Center for Atmospheric Research (NCEP for short) and the other is the European Centre for Medium-Range Weather Forecasts (ECMWF)'s most recent reanalysis denoted by ERA-40. Three analysis methods are considered, with increasing sophistication. Within each data set the analysis results are consistent with each other (i.e., each within the other's error bars), with the method of linear discriminant analysis (LDA) yielding the smallest error bar and the unfiltered global mean data yielding the largest error bar in the temperature amplitude. All three methods and both data sets are able to demonstrate that the 11-year signal is statistically significant and attributable (i.e., related) to the solar cycle. We deduce the spatial surface pattern over the globe which best distinguishes the solar maximum years from the solar minimum years using the LDA method. The resulting warming pattern shows clearly the polar amplification of warming and the preference for continents over oceans. We propose that the magnitude of the surface warming is consistent with direct solar radiative forcing if positive feedback processes such as ice albedo, water vapor/lapse rate and cloud feedbacks, similar to some of those studied for the greenhouse warming problem, are incorporated. It does not appear to be necessary to invoke some previously proposed exotic indirect mechanisms for an explanation of the observed solar signal.

Citation: Tung, K. K., and C. D. Camp (2008), Solar cycle warming at the Earth's surface in NCEP and ERA-40 data: A linear discriminant analysis, *J. Geophys. Res.*, *113*, D05114, doi:10.1029/2007JD009164.

1. Introduction

[2] Despite its name, the "solar constant" is not really invariant. It measures the amount of the total solar irradiance (TSI) impinging on a circular disk of unit area at the mean Sun-Earth radius from the Sun and varies on a number of timescales, both periodic and secular. We shall not be concerned here with the long-term trends of the Sun's radiation, but will concentrate only on the 11-year sunspot cycle, which will simply be called the solar cycle. Since November 1978, orbiting satellites have been able to measure the variation of the Sun's radiation from solar minimum (solar min) to solar maximum (solar max) for several cycles with good relative accuracy, and found that the TSI varies by approximately 1 W m^{-2} from solar min to solar max during the last three solar cycles [*Lean*, 1987; *Wilson et al.*, 1981], but appears to be weaker in the cycle in the 60s and early 70s. See, e.g., the black TSI curve in Figure 2c in section 2. The secular trends between cycles were not reliably measured because of intersatellite calibration problems. Using sunspot and other proxies, the TSI variations can be extended back a few hundred years prior to 1978, but we are interested only in the more accurate reconstruction since 1950s. In Lean's reconstruction [*Lean et al.*, 1995], there is very little secular trend in the TSI time series since 1950s. There is still some debate about this trend; we are interested only in the purely oscillatory behavior of the solar cycle forcing and response and so the trends are removed before the analysis.

[3] The Earth's climate response to the measured solar cycle forcing is also controversial because the expected surface warming is small and is imbedded in other atmospheric oscillations and noise. However, good instrumental measurement record is now available that span five solar cycles, and is probably long enough that with more sophisticated data analysis methods, the solar cycle signal can be extracted from the noise and its statistical significance established.

[4] There have been thousands of reports over 200 years of regional climate responses to the 11-year variations of solar radiation, ranging from cycles of Nile River flows,

[1]Department of Applied Mathematics, University of Washington, Seattle, Washington, USA.
[2]Now at Department of Mathematics, California Polytechnic State University, San Luis Obispo, California, USA.

0148-0227/08/2007JD009164$09.00

African droughts, to temperature measurements at various selected stations, but a coherent global signal at the surface has not yet been established statistically [*Hoyt and Schatten*, 1997; *Pittock*, 1978]. Since the forcing is global, theoretically one should expect a global-scale response. The simple method of global averaging turns out to be quite effective in filtering out the large unforced variability. When globally and annually averaged and detrended, but otherwise unprocessed, the NCEP surface air temperature [*Kalnay et al.*, 1996] since 1959 is seen in Figure 1 (reproduced from *Camp and Tung* [2007c]) to have an interannual variation of about 0.2–0.4 K, not exclusively of solar origin but nevertheless somewhat positively correlated with the solar cycle (the correlation coefficient ($\rho = 0.48$) turns out to be statistically significant at above 95% confidence level using an isospectral Monte Carlo test). The signal also contains a higher-frequency variation of comparable magnitude, possibly due to El Niño–Southern Oscillation (ENSO), and cooling after the volcanic eruptions of Pinatubo and El Chichón. The expected cooling following the eruption of El Chichón in 1982 and Pinatubo in 1991, both near the end the solar maxima, and the warming of the ENSO event of 1998, near the end of the solar minimum, together may give the visual appearance of shifting the temperature response ahead of the solar cycle forcing, yielding in some analyses the puzzling result that the response may lead the forcing by 1 to 2 years. See, for example, the time series analysis of *Coughlin and Tung* [2004]. There is no proposed mechanism for such a behavior. The extraction of a cleaner solar cycle response is desirable for a correct understanding and explanation.

[5] To filter out the nondecadal variability, we take advantage of the spatial characteristics of the solar cycle response. One rudimentary way of obtaining the spatial pattern objectively is to use the difference between the solar max composite and the solar min composite. This Composite Mean Difference (CMD) Projection method has been discussed by *Camp and Tung* [2007c]. Projecting the original detrended, annual mean data onto this spatial pattern yields a time series with the higher-frequency variability filtered out, yielding a higher correlation coefficient of $\rho = 0.64$, which is statistically significant, as judged by a bootstrap Monte Carlo test. The filtered response amplitude of $\kappa = 0.18 \pm 0.08$ K of global warming per W m^{-2} variation of solar constant TSI (not shown) is the same as that already apparent in the unfiltered global mean data in Figure 1, though with a 20% smaller error bar. We can do still better in reducing the error bar, using a more sophisticated optimization method. This is a subject of the present work.

2. Spatial-Time Filter

[6] Early estimates of the solar cycle response were obtained using (energy balance) model-generated "optimal space-time filter" [*Stevens and North*, 1996], whose pattern is small over the poles as compared to the tropics. We use here the method of linear discriminant analysis (LDA) developed by *Schneider and Held* [2001] originally to deduce the temperature trends, and later by *Camp and Tung* [2007a, 2007b] for studying the QBO, solar cycle and ENSO perturbations; more detail on the implementation of the method for the present problem, including mathematical formulae, can be found in the latter references. Although less intuitive than the CMD Projection method, the LDA method is better at filtering out other variability; we will show that the error bars from the LDA analysis is half as large as those obtained by the CMD method of *Camp and Tung* [2007a]. The input information used in the LDA method to construct the "solar cycle filter" is rather minimal and objective: it simply specifies what years are in the solar max group and what years belong to the solar min group. The LDA procedure, which maximizes the ratio R of the between-group variance relative to the variance within each group, then produces, via an optimization algorithm, the latitudinal weights from which we obtain both the filtered time series and the associated spatial pattern that best distinguish the solar max group from the solar min group by filtering out other atmospheric variability, such as ENSO. Previously used methods, multiple regressions and composite differences, have not been able to establish a statistically significant coherent global pattern although some were successful in finding regions of local significance; these methods by and large do not take advantage of the spatial information of the response. There is a subtle but important difference in the LDA approach used here as compared to methods that project the data onto a spatial pattern, including the EOF projection and the CMD projection [*Camp and Tung*, 2007c]: Using the present solar cycle signal problem as an example, the residual's spatial pattern obtained by the projection methods is orthogonal to the retained spatial pattern, but the residual can still contain in its time domain decadal (namely, 11-year period) signal. The residual in the LDA method, on the other hand, is orthogonal in the time domain to the retained (solar cycle) signal. Thus the solar cycle signal is optimally extracted in the retained mode.

[7] Figure 2a shows the meridional pattern objectively obtained for the zonal mean, annual mean air temperature at the surface using the global data set of NCEP, linearly detrended to remove the secular global warming signal. Figure 3a shows the corresponding temperature pattern in the 850–500 hPa layer, representing the lower troposphere. The surface pattern in Figure 2a, which is very close to the CMD pattern obtained previously, shows clearly the polar amplification of warming, predicted also by models for the global warming problem, with largest warming in the Arctic (3 times that of the global mean), followed by that of the Antarctic (2 times). There then must be an implied poleward heat transport. That the tropical lower atmosphere does not simply warm up to radiate to space the excess energy it receives is probably due to the fact that the warming and the implied horizontal temperature gradient that would have resulted would make the atmosphere more unstable to baroclinic instability, which leads to enhanced eddy poleward heat transport. The excess energy is transported by dynamic heat fluxes to the high latitudes, resulting in polar warming [*Cai*, 2005, 2006; *Cai and Lu*, 2007]. The poleward transport occurs rather quickly, probably within a year (and in any case less than 5 years), and probably involves mostly the atmosphere and the upper oceans, as *White et al.* [1997] showed that the solar cycle response does not penetrate deep enough into the ocean to engage the deep water. Low warming occurs over the latitudes of the

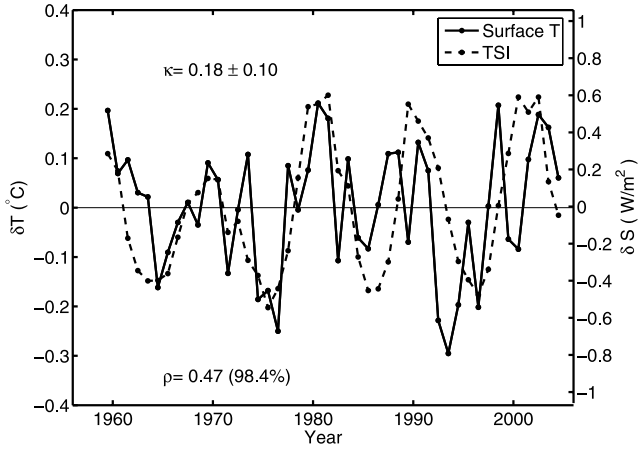

Figure 1. Annual mean, global mean NCEP surface air temperature (1959–2004), in solid line, with scale on the left axis. The dashed line shows the annual mean TSI time series [*Lean et al.*, 1995], updated and provided to us by J. Lean, with scale on the right axis. κ is the regression of global mean temperature response in K per each W m^{-2} variation of the solar constant. ρ is the correlation coefficient between the global temperature and the TSI. An isospectral Monte Carlo test, in which the spectral phase of the temperature (or the TSI) time series is randomized while preserving the spectral amplitude (so that the number of degrees of freedom of the observed time series is preserved) to generate 3000 synthetic time series, shows that this positive value of ρ is not likely to occur by chance, at 98.4% confidence level.

southern ocean and over the southern tropics. In general, warming over the oceans is much less than over land (see later). Over the tropics, no large warming occurs whether it is over land or over ocean. The warming over the tropics instead occurs higher up, at 200 hPa (not shown, at only 90% confidence level because of the quality of the upper air data prior to 1979), which is where the latent heat due to vertical convection is deposited. *Cai* [2005] discusses how the vertical transport of surface heating in a moist atmosphere leads to an increase in poleward heat transport despite the weakening of the surface temperature gradient. Many of the general features are similar to those predicted for global warming [*Manabe and Stouffer*, 1980].

[8] Using a bootstrap Monte Carlo test with replacement in Figures 2b and 3b, we show that a single optimal filter exists that separates the solar max years from the solar min years in temperature and that the large observed separability measure R could not have been obtained by chance at over 95% confidence level.

[9] Volcanic eruptions, particularly El Chichón in March 1982 and Pinatubo in June 1991, coincidentally occurring during solar maxima, may contaminate the 11-year signal. The expected cooling in the troposphere for the transient aerosol events however lasted temporarily, for about 2 to 3 years. Since the LDA analysis does not require a continuous time series, the volcano aerosol years can be excluded from the time series and a new discriminant pattern generated. This ensures that the discriminant pattern for solar cycle is not contaminated by the volcano response. This has been done in Figures 2 and 3, where the years 1982 and 1983 (after El Chichón), and 1992 and 1993 (after Pinatubo) are excluded. Removing a third year, or removing only 1 year, does not change the results reported. When no volcanic years were excluded in the LDA analysis, the warming amplitude is still the same but the confidence level is 4–5% lower (not shown).

[10] The projection of the annual means of NCEP reanalysis data for years from 1959 to 2004 onto the discriminant spatial weights is shown in Figure 2c and 3c. Given that our method requires only that the data be divided into two groups with no information on the peak amplitudes of either the solar irradiation or the temperature response, it is remarkable that the deduced global temperature response follows the solar radiation variability so well. The correlation coefficient, ρ = 0.84 and 0.85 in Figures 2 and 3, respectively, is highly statistically significant, higher than that obtained previously with the CMD projection method, although this method tends to yield high correlation coefficients in the filtered data and so the statistical tests have to be done starting with the original data before the two groups are defined. These tests establish that the surface (and lower tropospheric) temperature response is related to the solar cycle forcing at over 95% confidence level.

[11] Figure 2c visually reveals a global mean warming of slightly more than 0.2 K at the surface from solar min to solar max especially in the last three cycles. More conservatively, we fit $\delta T = \kappa \delta S$ to all 4.5 solar cycles, where $\delta S(t)$ is the TSI variability time series, and find $\kappa = 0.167 \pm 0.037$ K/(W m^{-2}) at the surface (and 0.213 ± 0.044 in 500–850 hPa). It should be noted that this measure of amplitude is not the peak-to-peak amplitude one commonly refers to, which can be obtained visually from Figure 2c to be about 0.2+ K, but is an average fit over the entire shape of the cycle. Consequently its amplitude is about 25% smaller than the peak amplitude. The error bars define a 95% confidence interval and are approximately equal to ±2 standard deviations (σ). This value of κ is about 70% higher than the regression coefficients of temperature against irradiance variability previously deduced [*Douglass and Clader*, 2002; *Lean*, 2005; *Scafetta and West*, 2005], of ~0.1 K global mean surface warming attributable to the solar cycles. Our higher response level is however consistent with some other recent reports [*Haigh*, 2003; *Labitzke et al.*, 2002; *Van Loon et al.*, 2004].

3. Error Analysis

[12] The error bar in κ shown above is due only to regression error. To see if there are other possible errors that give a larger error bar, we perform the so-called $N-1$ error analysis, in which we sequentially drop each year and perform a new LDA analysis until all possibilities are covered. This leads to $\kappa = 0.167 \pm 0.014$ at the surface (and 0.213 ± 0.020 in 500–850 hPa). The 2σ error bar is much smaller than the regression error, showing that the amplitude of κ is not affected by any one anomalous data point. Dropping m data points, if they are independent, increases the error bar relative to dropping one point by a factor of $m^{1/2}$. Monte Carlo simulations show that this is approximately true even without the independence assumption, for m not too large. The error bars from the N-m test would still be less than the regression error unless more than

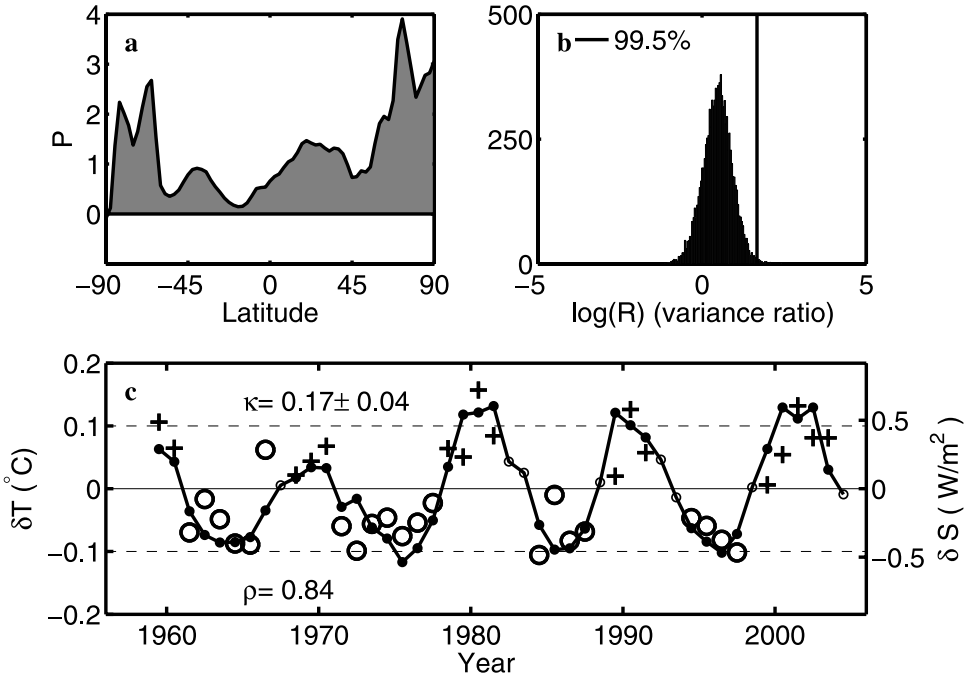

Figure 2. Surface temperature from NCEP 1959–2004. (a) The coherent latitudinal pattern which best distinguishes the years in the solar max group (when TSI is 0.06 W m^{-2} above the mean) from the years in the solar min group (when TSI is 0.06 W m^{-2} below the mean), normalized so that its global mean is one. (b) Bootstrap with replacement Monte Carlo test, showing that the separation R achieved by the pattern in Figure 2a, indicated by the vertical straight line, is not likely to be achieved by 10,000 time series generated by randomly assigning, with replacement, the same number of years to the solar max/min group as in the real data. (c) LDA filtered (projected) time series of temperature data. This projection is scaled such that the left axis shows the global mean temperature anomaly. To obtain the temperature anomaly at a particular latitude, multiply Figure 2a into Figure 2c. The pluses are temperatures in the solar max group, and the circles are in the solar min group. The black line shows the annual mean TSI time series with scale on the right axis. The small solid circles indicate the years used in the analysis, while the small open circles indicate the years dropped. These are the years of the volcanoes discussed in the text, and the years when the TSI variability is close to its mean, which are considered to be neither solar max nor solar min. Prior to the LDA analysis, NCEP time series at different latitudes are detrended and regularized (smoothed in space) using truncated SVD decomposition, at truncation level r = 17, chosen as discussed by *Camp and Tung* [2007a]. There is a range of truncation level for which we obtain over 95% confidence level. The particular confidence level shown in Figure 2b is specific to r = 17. We claim "over 95% confidence level," not 99.5%.

20% of the data are in error and dropped, which is unlikely. Thus, we obtain the following overall bounds for κ: κ = 0.17 ± 0.04 K/(W m^{-2}) for the surface air temperature response to variations in the solar constant.

[13] In NCEP reanalysis, temperature product is influenced by the model used in the reanalysis at the surface more than at constant pressure surfaces. We repeated the LDA analysis on the 925-hPa NCEP temperature, a "type A" product not much affected by model reanalysis, and obtained the same κ = 0.17 ± 0.04 K/(W m^{-2}), at well above 95% confidence level. See Figure 4.

[14] Instrumental errors are not included in our error bars. Because satellite measurement was not available until after 1978, our use of reconstructed TSI for the period 1959–1978 presents another source of error. An upper bound on this error is obtained by redoing the LDA dropping all years prior to 1979. We find that κ is reduced by 3%, a magnitude of difference well below the stated error bar. Note that the contamination of the signal by other variability, such as volcanoes and ENSO, has been minimized by our method. The greenhouse warming signal is removed to the extent possible by the linear trend. However, the linear trend may be sensitive to the end point and unfortunately 2005 is a very unusual year (one of the warmest on record). To minimize this end point error, only 1959–2004 were used in the analysis. To include 2005, a nonlinear trend may need to be used.

4. Detailed Spatial Pattern

[15] Having established the existence of a global-scale solar cycle response we can also examine in more detail the surface warming pattern over the globe. We repeat the LDA analysis on the gridded NCEP surface air temperature data at a latitude-longitude resolution of 5° × 5°. Consistent with the zonal mean pattern shown in Figure 2, the largest warming in Figure 5 occurs over the two Polar Regions.

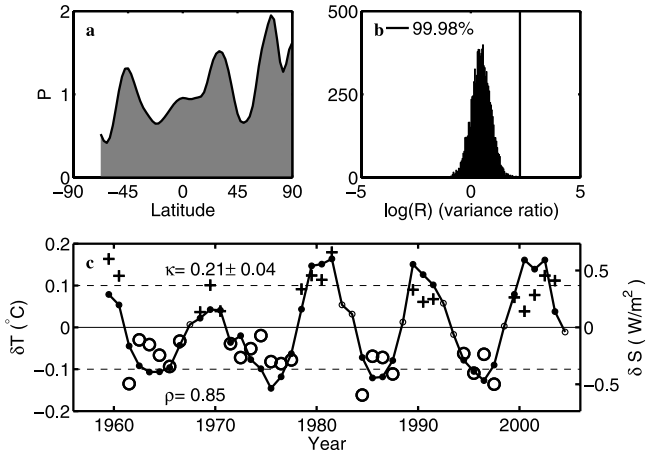

Figure 3. Same as in Figure 2 except for the mean temperature in the 850–500 hPa layer. Because the topography of the Antarctic continent protrudes into this layer even in zonal mean, the region 70–90°S is excluded. This exclusion affects the global mean temperature only minimally because of the small polar area.

Warming of close to 1 K occurs near seasonal sea ice edges in the Arctic Ocean. The largest warming occurs around the "Northwest Passage" (the Canadian Archipelago, Beaufort Sea, the coast of northern Alaska and the Chukchi Sea between Alaska and Siberia). The warm pattern is quite similar to the observed recent trend [*Moritz et al.*, 2002], and may suggest a common mechanism. A second polar warming center is located around the Antarctic continent on the seaward side, also suggestive of a positive sea ice albedo feedback as a mechanism for the polar amplification of the radiative forcing. (A cautionary note: Most of the stations in Antarctica are situated at the coast of that continent since 1957 with very few stations inland. This may have affected the spatial pattern obtained inland [*Uppala et al.*, 2005]. On the other hand, there are more inland stations in the Arctic and so the spatial pattern there may be more believable.) In the midlatitudes, there is more warming over the continents than over the oceans. Most of Europe is warmed by 0.5 K, and eastern Canada by 0.7 K, while western U.S. sees a smaller warming of 0.4–0.5 K. Iraq, Iran and Pakistan are warmer by 0.7 K and northern Africa by 0.5 K. Curiously the Andes in the South American continent is colder by 0.7 K.

[16] To ascertain the robustness of these patterns to whether the end of the time series occurs during a solar max or a solar min, the time series is truncated after the maximum of the last solar cycle in 2003 and again after the solar min of 1997, and the LDA repeated. The patterns in Figure 5 remain unchanged except that the Arctic warming gradually loses its detail with shorter and shorter records and becomes defused over the whole western half of the Arctic.

5. ERA-40 Data

[17] Since the LDA analysis requires continuous spatial information, any missing data would need to be filled in by some algorithm. We choose in this study to examine only the completed data provided by reanalysis. There are two available. The NCEP reanalysis is studied above. Next we compare it with the 45-year reanalysis of ERA-40, a second-generation reanalysis carried out by the European Centre for Medium-Range Weather Forecasts (ECMWF) [*Uppala et al.*, 2005]. It should be kept in mind the different way the surface temperature is derived in the two reanalyses. In ERA-40, the surface temperature is called the 2-m temperature. It is not obtained directly as part of the three-dimensional variational analysis of atmospheric fields, but is an interpolation from the lowest model level (at ~10 m) and the background forecast of the skin temperature. In NCEP reanalysis, the surface air temperature was derived from observations of upper air variables and surface pressure. *Simmons et al.* [2004] compared ERA-40, NCEP and Climate Research Unit (CRU) data directly derived from monthly station data, and found that the surface temperatures are in good agreement over the Northern and Southern Hemispheres, in Europe and in North America. However, because of the natural compensation of temperature anomalies between hemispheres, the differences in the global means between the two data sets may be larger as a percentage of the anomaly, as we are finding. In Figure 6, we compare the global mean annual mean variations in surface temperature from the two reanalyzed data sets. Three noticeable differences are (1) the ERA-40 global temperature is offset by 0.5 K, hotter than the NCEP data; (2) there is a positive linear trend in the NCEP data, but in the ERA-40 data, the linear trend is negative prior to 1977 and positive after that year; and (3) the interannual variability of the global mean temperature is about a quarter smaller in the ERA-40 data. *Gleisner et al.* [2005] also found differences between the two reanalysis data sets in the tropospheric temperature above the surface and attributed them to likely satellite correction biases in the ERA-40 data resulting in a weaker solar cycle signal.

[18] The LDA method is not affected at all by the offset. It is slightly affected by how the trend is taken; the CMD Projection method is more sensitive because it depends on a difference of the two groups defined after the trend is taken, while the LDA method maximizes the difference between the two groups even in the presence of a residual trend. The

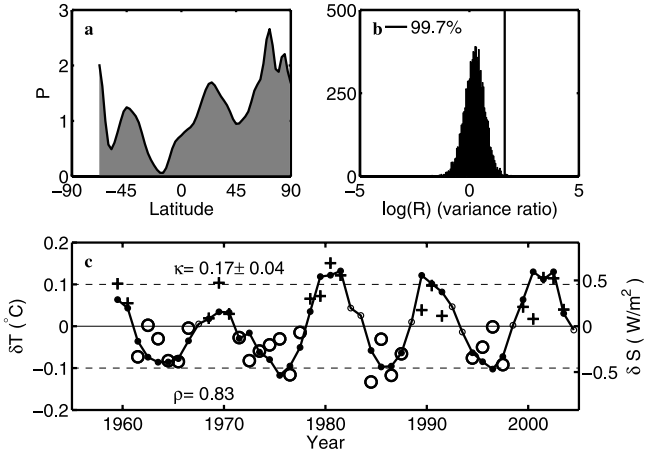

Figure 4. Same as Figure 2 except using the 925-hPa NCEP data.

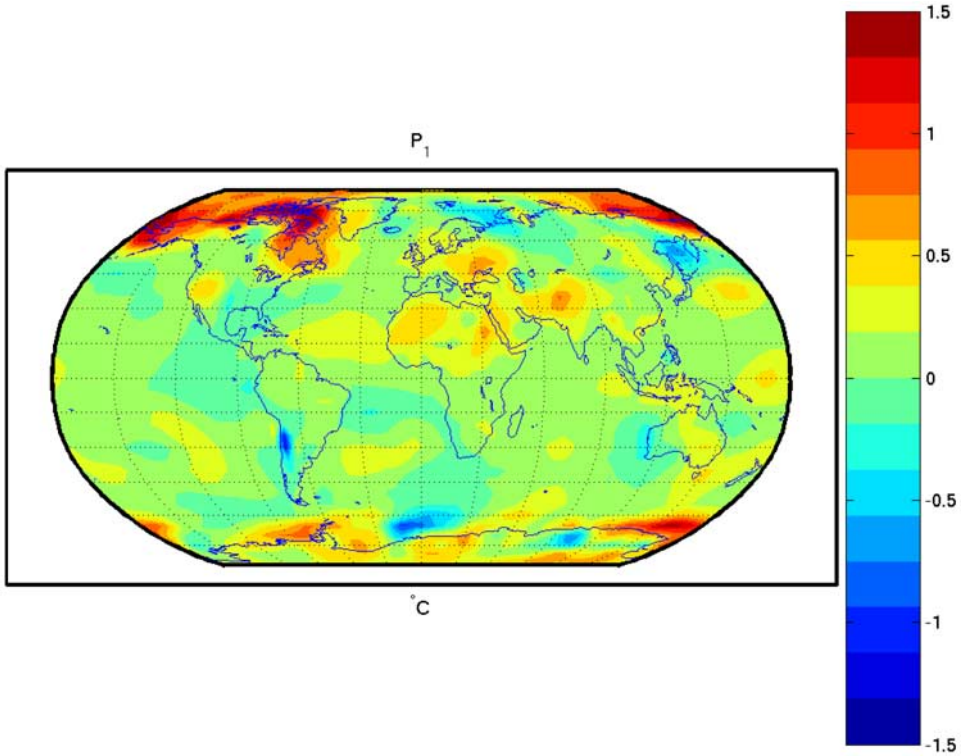

Figure 5. The global surface pattern of temperature that best distinguishes the solar max group from the solar min group. Shown in color is the temperature difference in K between ±one standard deviation from the mean. The actual peak-to-peak difference between the solar max and solar min is larger but is not as robust as the standard deviation difference. A measure of the peak-to-peak difference can be obtained by multiplying the values shown by a factor of $\pi/2$. Monte Carlo test shows that this global pattern is statistically significant above the 95% confidence level.

last difference, that the variance of the ERA-40 data is smaller than that of the NCEP data, affects the magnitude of the regression coefficient κ obtained by either method but does not affect the statistical significance of the existence of the solar cycle signal. That is, the existence of a globally coherent pattern of warming by the solar cycle is robustly determined by the LDA and the CMD methods at over 95% confidence level using either data set.

[19] Figure 7, shows the global mean annual mean ERA-40 surface temperature, linearly detrended for the period 1958–2002. Its correlation coefficient with the TSI index is $\rho = 0.41$, about same as that for the NCEP data, and is also statistically significant at over 95% confidence level under the isospectral Monte Carlo test. Its regression (against the TSI index) amplitude is $\kappa = 0.14 \pm 0.10$ and is bounded away from zero on the positive side, also at over 95% level.

[20] Figure 8 shows the 2-D LDA analysis of the global ERA-40 data surface temperature. In general the patterns is very close to that shown in Figure 5 for the NCEP surface temperature, although the overall variance of the signal is somewhat smaller in the ERA-40 case. It shows the polar amplification of warming, with 3 times the global mean over the Arctic and twice the global mean over the Antarctic, and the general preference of warming over continents than over the oceans at midlatitudes. The centers of warming are the same in the two data sets. The difference appears to be related to the strength of negative anomalies. Over the Antarctica continent, there is more of a negative temperature anomaly in the ERA-40 data than in the NCEP data. This difference may not be meaningful given the sparseness of measuring stations there. Over the eastern Arctic, the negative anomaly in the ERA-40 result appears stronger than in the case of NCEP. The LDA-filtered data shows a very high correlation coefficient with the solar TSI, $\rho =$

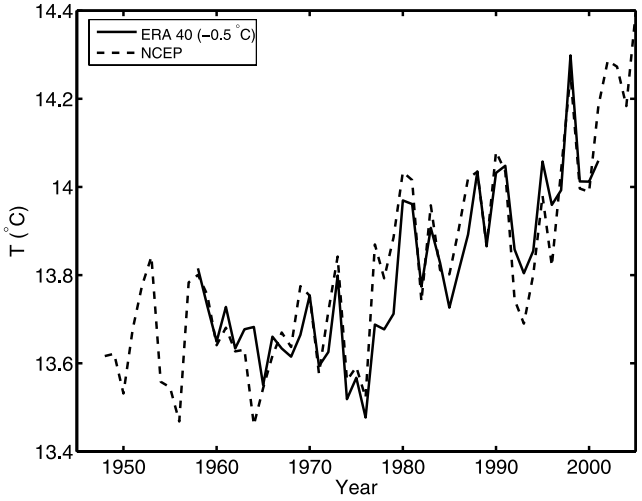

Figure 6. Comparison of the global mean annual mean surface temperature in ERA-40 offset by 0.5°C (thick solid line) and in NCEP (dashed line).

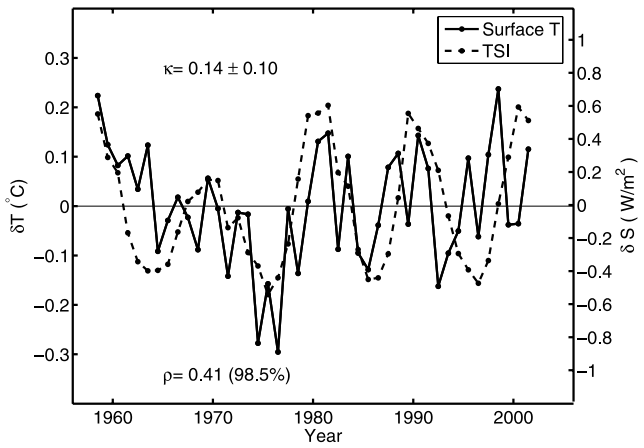

Figure 7. Same as Figure 1 except using ERA-40 data, linearly detrended for the period 1958–2002.

0.84, same as in the NCEP case. The regression amplitude of the 2D-LDA-filtered data is $\kappa = 0.13 \pm 0.03$ K per each W m^{-2} of variation of the solar constant, which is smaller than in the NCEP case, though still near its lower range. The peak-to-peak amplitude of the global temperature response is still close to 0.2 K for the last three solar cycles.

[21] Given the fact that the surface temperature is derived in different ways in the two re analyses, perhaps it is more meaningful to compare the results for the 925-hPa level. The LDA analysis for the zonal mean 925-hPa ERA-40 temperature is shown in Figure 9. This is to be compared with the NCEP result shown in Figure 4. The correlation coefficient with the solar TSI is $\rho = 0.83$, same as in the NCEP case and both are highly statistically significant. The regression amplitude is $\kappa = 0.14 \pm 0.03$ K/W m^{-2}, which is smaller than the corresponding value in the NCEP result, but within its lower bound. The peak-to-peak amplitude determined visually in Figure 9 is still close to 0.2 K for the last three solar cycles, with that in the NCEP data being slightly higher, and that in the ERA-40 just about 0.2 K. Despite these differences in amplitude, the existence of a global solar cycle response is robustly established in both data sets and using three different methods.

6. Explaining the Solar Cycle Response

[22] Although previously attention has been focused on the UV part of the solar cycle and its absorption by ozone in the stratosphere, the amount of the total solar irradiance (TSI) reaching the Earth's surface is not negligible. The observed ~1.0 W m^{-2} variation of the solar constant from solar min to solar max in the last three solar cycles translates into a net radiative heating of the lower troposphere (mostly the surface) of $\delta Q = \frac{1.0 \cdot 0.85}{4} \sim 0.2$ W m^{-2}. The factor of 4 is to account for the difference between a unit area on the spherical Earth and the circular disk on which the solar constant is measured, while 0.85 is to account for the 15% of the TSI variability that lies in the UV wavelength and is absorbed by ozone in the stratosphere with the remaining reaching the lower troposphere, the surface and the upper ocean [*Lean et al.*, 2005; *White et al.*, 1997]. That 15% of the solar radiation variability is not entirely lost to the lower atmosphere. *Shindell et al.* [2006] found that this ozone heating in the stratosphere exerts a small positive forcing of the troposphere of ~0.03 W m^{-2}. More detailed estimates can also be done. Assuming no change in the troposphere, *Larkin et al.* [2000] estimated a net radiative heating at the

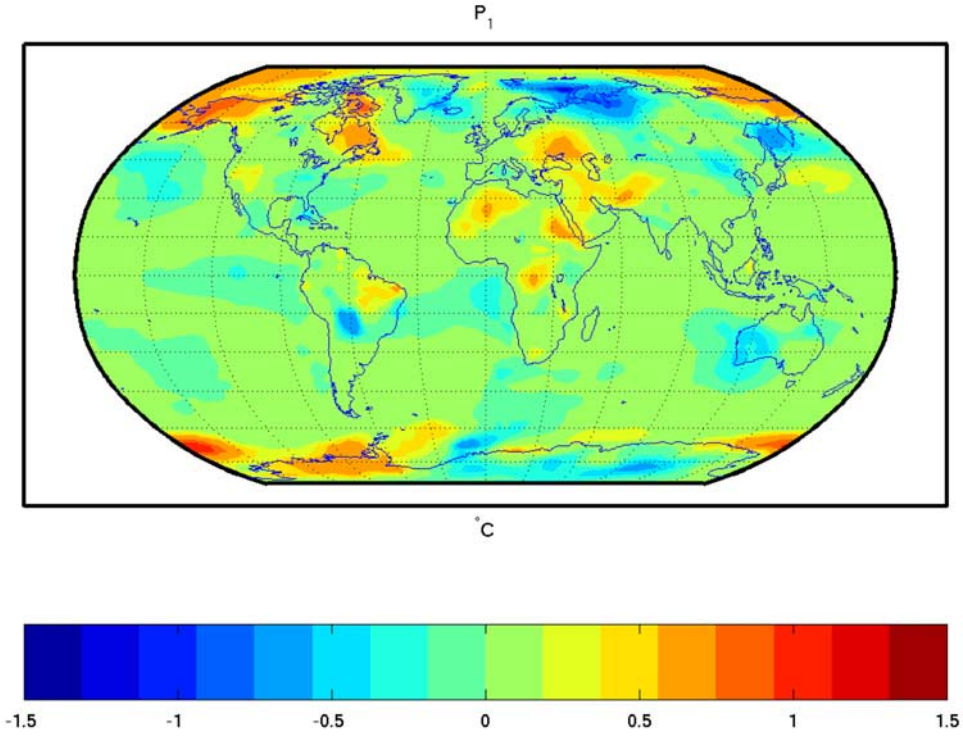

Figure 8. Same as in Figure 5 except using ERA-40 surface temperature.

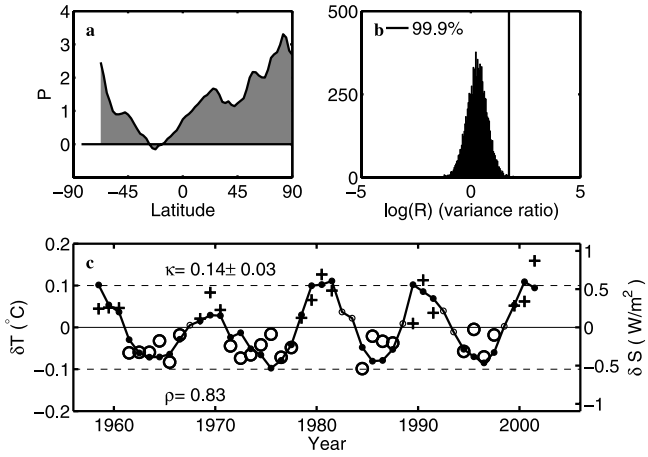

Figure 9. Same as Figure 4 except with 925-hPa ERA-40 temperature.

top of the troposphere of 0.18–0.26 W m^{-2} from solar min to solar max when no ozone change is taken into account. When the effect ozone increase due to the solar cycle is incorporated in the two GCM experiments, the authors found $\delta Q \sim 0.23$ W m^{-2}. In the work by *Forster et al.* [2007, chap. 2], the radiative forcing (RF) of the 11-year solar cycle variability is not reduced by the stratospheric absorption citing compensation by indirect effect of solar-ozone interaction in the stratosphere (see footnote 11 therein). The RF is given by the net radiative heating reduced by the portion reflected back to space:

$$RF = (1 - \alpha)\delta Q \sim 0.18 \text{ W m}^{-2},$$

where $\alpha \sim 0.3$ is the albedo, the fraction reflected by the surface and clouds.

[23] This solar radiative forcing is about 1/20 that for doubling CO_2 ($RF \sim 3.7$ W m^{-2}). Thus the annual rate of increase in radiative forcing of the lower atmosphere from solar min to solar max happens to be equivalent to that from a 1% per year increase in greenhouse gases, a rate commonly used in greenhouse gas emission scenarios [*Houghton et al.*, 2001].

[24] In the absence of fast feedbacks, the tropospheric RF of ~0.18 W m^{-2} from solar min to solar max is balanced by infrared reemission and it would have produced at equilibrium at the surface a temperature change of $\delta T \sim \delta Q\ (1 - \alpha)/B = RF/B \sim 0.1$ K. (The increase in infrared reemission is given by $B\delta T$ with $B = 1.9$ W m^{-2} per K [*Graves et al.*, 1993].) Our observed global mean warming of ~0.2 K would seem to imply that, if it is due to TSI heating at the surface, the fast feedback processes in our atmosphere, such as ice albedo, lapse rate, water vapor and cloud feedbacks, should in aggregate amplify the initial TSI warming by about a factor of $f \sim 2$–3. (One should take into account the fact that the phenomenon is periodic and not at equilibrium; see Appendix A.) From the large body of work on radiative feedback processes related to the global warming problem [*Bony et al.*, 2006], we know that a "climate amplification factor" of this range is justifiable physically. Because of the fast timescales involved in these processes, it is reasonable to expect that the same feedback factor applies to the decadal phenomenon as well. Previous GCM calculations [*Haigh*, 1996; *Shindell et al.*, 1999] have tended to underestimate the response to solar cycle forcing possibly because, as pointed out by *Haigh* [1996], the fixed sea surface temperature in these models might have reduced the surface heating and the magnitude of the feedback processes. A new GCM run with coupled dynamical ocean by *Shindell et al.* [2006], yielded a ~0.1 K globally average surface response as the difference between a perpetual solar max run and a perpetual solar min run. The authors however attribute most of the increase in response to interactive ozone chemistry.

[25] In the troposphere the phenomena of solar cycle and global warming are quite similar. The radiative forcing for both is global in extent and relatively uniform, although solar forcing occurs only where the Sun shines. (Our use of annual means aims at reducing this difference.) The main difference lies in the stratosphere, but the effect of these differences on the near surface temperature is expected to be small. The stratosphere in solar max warms because of ozone absorption of the UV portion of the solar constant variation, which, with a variability of 0.12 W m^{-2} [*Lean et al.*, 2005], is larger, in percentage terms, than the variability in the TSI. The effect of the solar cycle ozone warming in the tropical stratosphere, which is about 0.5–1.5 K, on the lower troposphere has been investigated by GCMs [*Haigh*, 1999; *Shindell et al.*, 1999] and is found to be small: *Haigh* [1996] found that the Hadley circulation is shifted slightly, by 0.7 degrees of latitude. There is evidence in our Figure 3a of the two midlatitude strips of warming suggested by her as a result of this shift. *Shindell et al.* [1999] found that on a global mean basis, the net surface warms by about 0.07 K, including both the stratospheric influence and direct heating of the surface (but with fixed sea surface temperature). The observed solar cycle related heating over the polar stratosphere is larger, at a mean of 4 K [*Camp and Tung*, 2007a], but this occurs only during late winter and over a small area, related to the enhanced frequency of occurrence of the Stratospheric Sudden Warming phenomenon [*Labitzke*, 1982]. Although the effect can be transmitted to the polar troposphere [*Baldwin and Dunkerton*, 1999], the anomaly near the surface on a global and annual mean is small. If these stratospheric differences can be ignored, the surface warming seen in Figure 2 in the zonal mean, and in more detail in Figure 5, may give a hint of the initial transient greenhouse warming at the surface in 5–6 years. This is because at a projected 1% increase per year of the greenhouse gases it takes about 5 years to increase the radiative forcing to the 0.18 W m^{-2} responsible for the response shown in Figures 2 and 4. Longer than a few decades, response to a monotonically increasing forcing in the greenhouse gas problem engages the deep water, and the two problems cannot be scaled.

7. Conclusion

[26] Using NCEP reanalysis data that span four and a half solar cycles, we have obtained the spatial pattern over the globe which best separates the solar max years from the solar min years, and established that this coherent global pattern is statistically significant using a Monte Carlo test. The pattern shows a global warming of the Earth's surface

of about 0.2 K, with larger warming over the polar regions than over the tropics, and larger over continents than over the oceans. It is also established that the global warming of the surface is related to the 11-year solar cycle, in particular to its TSI, at over 95% confidence level. The robustness of these results is tested with different periods and with different data products. In particular we have been able to confirm the existence of the solar cycle warming at the surface also in the ERA-40 data set, and obtain the same statistical significance. The amplitude of the regressed solar cycle signal in the ERA-40 data is, however, smaller, because of the overall smaller variance in the raw ERA-40 data as compared to the raw NCEP data. The peak-to-peak amplitude of the solar cycle response in the ERA-40 data is nevertheless still around 0.2 K.

[27] There have been many puzzling statements in the literature about the difficulty of explaining the solar cycle response at the surface given the magnitude of the solar forcing. Although the GCM simulation of a surface solar cycle response remains a challenge, we show here that given the magnitude of the measured amplitudes of forcing and response, the observed 0.2 K of warming is explainable assuming direct radiative forcing and the commonly accepted climate gain factor of $f \sim 2-3$, due to ice albedo, cloud and water vapor/lapse rate feedbacks.

Appendix A: Analysis—Energy Balance at the Surface

[28] Previous calculations using Energy Balance Models (EBMs) have yielded a response of 0.03 K to 0.06 K of warming to the solar cycle forcing, much smaller than even the previously extracted response of 0.1 K from observations. The perception that the response may not be energetically consistent with the radiative forcing has acted to encourage proposals to study indirect mechanisms of solar cycle response, some of which are sound while others are quite speculative, and downplay the primary importance of the mechanism of direct radiative forcing. The purpose of this section is to show that the observed solar cycle response is energetically consistent with the magnitude of the direct radiative forcing and with typical and reasonable values of ocean heat flux and atmospheric feedback amplifications. It is however not meant to be a model calculation of the solar cycle response.

[29] Consider the heat budget of the part of the atmosphere near the surface and the upper few meters of the ocean that directly absorbs solar heating and radiates back to space in infrared, where $T(y, t)$ is the surface temperature:

$$C\frac{\partial}{\partial t}\overline{T} = Q(1-\overline{\alpha}) - \left(A + B\overline{T}\right) + \frac{\partial}{\partial z}\overline{F_z}, \qquad \text{(A1)}$$

where the overhead bar denotes global averaging. Equation (A1) states that the heat content of this layer is increased by radiative forcing (first term on the right) and by heat flux from the oceans below (the last term), and decreased by infrared emission to space above (second term). The global average removes the meridional dynamical transport of heat term, since the latter is usually written in the form of a divergence. However, the presence of poleward heat transport and polar amplification of warming can increase the global mean warming by 10% [*Cai*, 2005]. This is ignored. Q is $\frac{1}{4}$ of the solar constant, and α(y) is the albedo: the fraction of the Sun's radiation reflected back to space by clouds and surface. $(A + BT)$ is the linearized form of the infrared emission of the Earth to space fitted from observational data on outgoing long-wave radiation, with $A = 202$ W m^{-2}, and $B = 1.90$ W m^{-2} K^{-1} in the current climate. They are temperature-dependent if the current climate is perturbed. The parameter C in equation (A1) represents the thermal capacity of the atmosphere (not of the deep ocean, as some authors assumed). We write $\tau = C/B$, which measures the timescale due to the atmosphere's thermal inertia. $\overline{\alpha}$ is the weighted global average albedo. The overbar is henceforth dropped for convenience. Considering small radiative perturbation δQ in $Q = Q_0 + \delta Q$, the equation governing the small temperature perturbation can be obtained from the first variation of the above equation, with B and α expanded in a Taylor series in T. This leads to the following perturbation equation:

$$B\tau\frac{\partial}{\partial t}\delta T = (1-\alpha)\delta Q - B\delta T/f + \frac{\partial}{\partial z}\delta F_z,$$

where

$$f = 1/(1-g),$$
$$g = \left(-\frac{\partial}{B\partial T}\,A - \frac{T}{B}\frac{\partial}{\partial T}\,B - \frac{Q}{B}\frac{\partial}{\partial T}\alpha\right)_0 \qquad \text{(A2)}$$

The factor f is the controversial climate gain, and g is the effect of temperature-dependent feedback factors, including the water vapor feedback (in the first two terms) and ice and snow albedo feedback (in the third term). Cloud feedback has contributions in all three terms. Some previous EBMs do not allow the radiative parameters such as albedo to change in a warmed climate, and consequently the factor f is missing. Some other EBMs reinsert in an ad hoc fashion tunable parameters that play the role of f, but in the wrong part of the equation.

[30] For solar cycle response, we model the flux to the ocean as diffusive (i.e., $F_z = -CD\partial T/\partial z$) with an exponential decay scale in the upper ocean as if it is semi-infinite (and so $\partial(\delta T)/\partial z = -\mu\delta T$). This is equivalent to neglecting the main thermocline; this is appropriate for the solar cycle response, which does not penetrate deep enough into the ocean. Thus the last term in (4) becomes $-CD\mu^2\delta T$, which is a loss term in the energy balance of the surface. Periodic solution:

$$\text{If}: \delta Q = a\cos(\omega t),$$
$$\text{then}: \delta T = \frac{(1-\alpha)\delta Q(t-\Delta)\widetilde{f}}{B}\,\frac{1}{\sqrt{1+\varepsilon^2}}, \qquad \text{(A3)}$$
$$\text{where}: \varepsilon = \widetilde{f}\omega\tau; \omega\Delta = \tan^{-1}(\varepsilon);\ \widetilde{f} \equiv \frac{f}{1+D\mu^2 f\tau}.$$

Compared to the steady state solution for a steady forcing, the periodic solution is delayed by the phase lag of Δ, and its amplitude is diminished by the factor $(1 + \varepsilon^2)^{-1/2}$, and further diminished by the diffusive heat loss to the oceans, by the factor $1/(1 + D\mu^2 f\tau)$. Since the phase lag and the amplitude factor are related, an observation of the phase lag

of the solar cycle also gives an estimate of the amplitude ratio between the periodic solution and the equilibrium solution.

[31] For an oscillating heating which reverses every 5.5 years, we do not expect the solar cycle heating to penetrate too deeply into the ocean. *White et al.* [1997] found that the solar cycle signal penetrated only $1/\mu \sim 100$ m into the upper ocean, with no effect from the deep water below the main thermocline, and that the observed phase lag in the ocean response peaks at zero lag but with an error bar of 2 years. Atmospheric lag should be shorter than the lag in the ocean response. The correlation coefficient ρ between the atmospheric temperature projection and the solar flux also peaks at zero phase lag and drops for larger lags, except possibly for a lag or lead of 1 year (separate LDA analysis with shifted time series not shown). If the observed phase lag Δ is indeed $\sim$1 year, we find $(1+\varepsilon^2)^{-1/2} \sim 1/1.19$. For a typical value of ocean diffusivity of $D \sim 1.0$ cm^2/s, we can back-deduce that the atmosphere's thermal inertia timescale is given by $f\tau = 1/[\omega/\varepsilon - D\mu^2] \sim 1.73$ year.

[32] Since our observational analysis uses annual mean data, we do not have an accurate determination of the phase lag of the solar response. We think it is less than 1 year for the reasons given above. We shall bracket the response using zero and 1 year lag. Equation (A3) then yields, for $f = 2$:

$$\delta T = \frac{(1-\alpha)\widetilde{f}}{B\sqrt{1+\varepsilon^2}}\delta Q \sim 0.19 \text{ K}/\left(\text{W m}^{-2}\right) \text{ for no phase lag.}$$

For the possible case of a lag of $\Delta \sim 1$ year, we will need to have to a larger $f = 3.2$ to get the observed (NCEP) temperature response of 0.17 K. Thus we consider the global surface temperature response of to the 11-year solar cycle explainable primarily by TSI forcing magnified by a factor of $f \sim 2$–3 climate gain due to the fast feedback processes. This same f should possibly also apply to the fast climate gain factors involved in greenhouse gas radiative heating.

[33] **Acknowledgments.** The research is supported by grant ATM-03 32364 from National Science Foundation, Climate Dynamics Program. We thank Judith Lean for providing us with her reconstructed TSI and UV time series.

References

Baldwin, M. P., and T. J. Dunkerton (1999), Propagation of the Arctic Oscillation from the stratosphere to the troposphere, *J. Geophys. Res.*, *104*, 30,937–30,946.

Bony, S., et al. (2006), How well do we understand and evaluate climate change feedback processes, *J. Clim.*, *19*, 3445–3482.

Cai, M. (2005), Dynamical amplification of polar warming, *Geophys. Res. Lett.*, *32*, L22710, doi:10.1029/2005GL024481.

Cai, M. (2006), Dynamical greenhouse-plus feedback and polar warming amplification. Part I: A dry radiative-transportive climate model, *Clim. Dyn.*, *26*, 661–674, doi:10.1007/s00382-005-010406.

Cai, M., and J. Lu (2007), Dynamical greenhouse-plus feedback and polar warming amplification. Part II: Meridional and vertical asymmetries of the global warming, *Clim. Dyn.*, *29*, 375–391, doi:10.1007/s00382-007-0238-9.

Camp, C. D., and K. K. Tung (2007a), The influence of the solar cycle and QBO on the late winter stratospheric polar vortex, *J. Atmos. Sci.*, *64*, 1267–1283.

Camp, C. D., and K. K. Tung (2007b), Stratospheric polar warming by ENSO in winter: A statistical study, *Geophys. Research. Lett.*, *34*, L04809, doi:04810.01029/02006GL028521.

Camp, C. D., and K. K. Tung (2007c), Surface warming by the solar cycle as revealed by the composite mean difference projection, *Geophys. Res. Lett.*, *34*, L14703, doi:10.1029/2007GL030207.

Coughlin, K., and K. K. Tung (2004), Eleven-year solar cycle signal throughout the lower atmosphere, *J. Geophys. Res.*, *109*, D21105, doi:10.1029/2004JD004873.

Douglass, D. H., and B. D. Clader (2002), Climate sensitivity of the Earth to solar irradiance, *Geophys. Res. Lett.*, *29*(16), 1786, doi:10.1029/2002GL015345.

Forster, P. M., et al. (2007), Changes in atmospheric constituents and in radiative forcing, in *Climate Change 2007: The Physical Science Basis—Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by S. Solomon et al., pp. 129–234, Cambridge Univ. Press, New York.

Gleisner, H., P. Thejll, M. Stendel, E. Kaas, and B. Machenhauer (2005), Solar signals in tropospheric re-analysis data: Comparing NCEP/NCAR and ERA-40, *J. Atmos. Sol. Terr. Phys.*, *67*, 785–791.

Graves, C. E., W. H. Lee, and G. R. North (1993), New parameterizations and sensitivities for simple climate models, *J. Geophys. Res.*, *98*, 5025–5036.

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

Haigh, J. D. (1999), A GCM study of climate change in response to the 11-year solar cycle, *Q. J. R. Meteorol. Soc.*, *125*, 871–892.

Haigh, J. D. (2003), The effects of solar variability on the Earth's climate, *Philos. Trans. R. Soc. London, Ser. A*, *361*, 95–111.

Houghton, J. T., Y. Ding, D. J. Griggs, M. Noguer, J. van der Linden, and D. Xiaosu (2001), *Climate Change 2001: The Scientific Basis*, Cambridge Univ. Press, New York.

Hoyt, D. V., and K. H. Schatten (1997), *The Role of the Sun in Climate Change*, 279 pp., Oxford Univ. Press, New York.

Kalnay, E., et al. (1996), The NCEP/NCAR 40-year reanalysis project, *Bull. Am. Meteorol. Soc.*, *77*, 437–471.

Labitzke, K. (1982), On the interannual variability of the middle stratosphere during the northern winters, *J. Meteorol. Soc. Jpn.*, *60*, 124–139.

Labitzke, K., J. Austin, N. Butchart, J. Knight, M. Takahashi, M. Nakamoto, T. Nagashima, J. Haigh, and V. Williams (2002), The global signal of the 11-year solar cycle in the stratosphere: Observations and models, *J. Atmos. Sol. Terr. Phys.*, *64*, 203–210.

Larkin, A., J. D. Haigh, and S. Djavidnia (2000), The effect of solar UV irradiance variations on the Earth's atmosphere, *Space Sci. Rev.*, *94*, 199–214.

Lean, J. (1987), Solar ultraviolet irradiance variations: A review, *J. Geophys. Res.*, *92*, 839–868.

Lean, J. (2005), Living with a variable sun, *Phys. Today*, *58*, 32–38.

Lean, J., J. Beer, and R. Bradley (1995), Reconstruction of solar irradiance since 1610: Implications for climate change, *Geophys. Res. Lett.*, *22*, 3195–3198.

Lean, J., G. J. Rottman, J. Harder, and G. Kopp (2005), SORCE contributions to new understanding of global change and solar variability, *Sol. Phys.*, *230*, 27–53.

Manabe, S., and R. J. Stouffer (1980), Sensitivity of a global climate model to an increase of CO_2 concentration in the atmosphere, *J. Geophys. Res.*, *85*, 5529–5554.

Moritz, R. E., C. M. Bitz, and E. J. Steig (2002), Dynamics of recent climate change in the Arctic, *Science*, *297*, 1497–1502.

Pittock, A. B. (1978), A critical look at long-term sun-weather relationships, *Rev. Geophys.*, *16*, 400–420.

Scafetta, N., and B. J. West (2005), Estimated solar contribution to the global surface warming using the ACRIM TSI satellite composite, *Geophys. Res. Lett.*, *32*, L18713, doi:10.1029/2005GL023849.

Schneider, T., and I. M. Held (2001), Discriminants of twentieth-century changes in earth surface temperatures, *J. Clim.*, *14*, 249–254.

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

Shindell, D. T., G. Faluvegi, R. L. Miller, G. A. Schmidt, J. E. Hansen, and S. Sun (2006), Solar and anthropogenic forcing of tropical hydrology, *Geophys. Res. Lett.*, *33*, L24706, doi:10.1029/2006GL027468.

Simmons, A. J., P. D. Jones, V. da Costa Bechtold, A. C. M. Beljaars, P. W. Kållberg, S. Saarinen, S. M. Uppala, P. Viterbo, and N. Wedi (2004), Comparison of trends and low-frequency variability in CRU, ERA-40, and NCEP/NCAR analyses of surface air temperature, *J. Geophys. Res.*, *109*, D24115, doi:10.1029/2004JD005306.

Stevens, M. J., and G. R. North (1996), Detection of the climate response to the solar cycle, *J. Atmos. Sci.*, *53*, 2594–2608.

Uppala, S. M., et al. (2005), The ERA-40 re-analysis, *Q. J. R. Meteorol. Soc.*, *131*, 2691–3012.

Van Loon, H., G. A. Meehl, and J. M. Ablaster (2004), A decadal solar effect in the tropics in July–August, *J. Atmos. Sol. Terr. Phys.*, *66*, 1767–1778.

White, W. B., J. Lean, D. R. Cayan, and M. D. Dettinger (1997), Response of global upper ocean temperature to changing solar irradiance, *J. Geophys. Res.*, *102*, 3255–3266.

Wilson, R. C., S. Gulkis, M. Janssen, H. S. Hudson, and G. A. Chapman (1981), Observation of solar irradiance variability, *Science*, *211*, 700–702.

C. D. Camp, Department of Mathematics, California Polytechnic State University, San Luis Obispo, CA 93407, USA.

K. K. Tung, Department of Applied Mathematics, University of Washington, Seattle, WA 98195, USA. (tung@amath.washington.edu)