Solar Total Irradiance Variations and the Global Sea Surface Temperature Record

George C. Reid

Aeronomy Laboratory, National Oceanic and Atmospheric Administration, Boulder, Colorado

The record of globally averaged sea surface temperature (SST) over the past 130 years shows a highly significant correlation with the envelope of the 11-year cycle of solar activity over the same period. This correlation could be explained by a variation in the sun’s total irradiance (the solar "constant") that is in phase with the solar-cycle envelope, supporting and updating an earlier conclusion by Eddy (1976) that such variations could have played a major role in climate change over the past millennium. Measurements of the total irradiance from spacecraft, rockets, and balloons over the past 25 years have provided evidence of long-term variations and have been used to develop a simple linear relationship between irradiance and the envelope of the sunspot cycle. This relationship has been used to force a one-dimensional model of the thermal structure of the ocean (Hoffert et al., 1980), consisting of a 100-m mixed layer coupled to a deep ocean and including a thermohaline circulation. The model was started in the mid-seventeenth century, at the time of the Maunder Minimum of solar activity, and mixed-layer temperatures were calculated at 6-month intervals up to the present. The total range of irradiance values during the period was about 1%, and the total range of SST was about 1°C. Cool periods, when temperatures were about 0.5°C below present-day values, were found in the early decades of both the nineteenth and twentieth centuries. There is direct evidence for the latter period from the historical SST record and some indirect evidence for the earlier cool period. While many aspects of the study are unavoidably simplistic, the results can be taken as indicating that solar variability has been an important contributor to global climate variations in recent decades. It has probably not been the only contributor, however, and in particular, the growing atmospheric burden of greenhouse gases may well have played an important role in the immediate past. This role is likely to become even more important in the near future.

1. INTRODUCTION

Global climate change is rapidly becoming one of the major scientific issues of the decade, largely because it has a potentially greater impact on world society than almost any other phenomenon, barring such catastrophes as nuclear war or collision with an asteroid, and because civilization itself may well be a root cause through anthropogenic changes in the composition of the atmosphere. Unfortunately, such changes have to be detected against the background of natural climate variability, whose nature and causes are still uncertain and controversial.

Since the sun’s radiation is the principal driving force for global climate, it is natural to suspect variations in solar irradiance as a possible source of climate variations. Speculations along these lines have abounded over the years and provided the rationale for the extended series of measurements of the solar constant carried out by the Smithsonian Astrophysical Observatory [Abbott, 1953] during the early decades of this century. While these measurements did show short-term variations, the possibility that they represented fluctuations in atmospheric transmission, or errors in measurement, rather than true variations in the sun’s irradiance could not be eliminated [Hoyt, 1979]. All ground-based measurements of the sun’s irradiance face the difficulty of variable atmospheric transmission, and no reliable information on irradiance variations was obtained until improved techniques and platforms located above the atmosphere began to appear on the scene in the 1960s [Fröhlich, 1977]. A decade-long series of solar irradiance measurements is now available from radiometers carried on board spacecraft, and has revealed variations on a wide range of time scales, including that of the 11-year cycle of solar activity [Wilson et al., 1986; Hickey et al., 1988; Hudson, 1988]. There is as yet no reliable information concerning variability on the longer time scales that are likely to be most effective in causing climate change, but presumably such information will eventually become available if the existing series of measurements is maintained.

On the climate side, the problem has been not so much a shortage of information as an overabundance. "Global" climate is an elusive concept and can only be defined as some kind of global average of a large number of localized regional climates, each of which can in turn be defined by a variety of parameters, such as temperature, precipitation, atmospheric pressure, and wind speed. Whether or not global climate is a meaningful concept is debatable, but if a single parameter has to be selected as an indicator of global climate, then temperature is probably the most reasonable one to choose, since the temperature of a body that is absorbing radiation is known to vary monotonically with the intensity of the radiation field, and most of the other climate parameters are functions of temperature directly, or of temperature gradients through their influence on atmospheric dynamics.

Measurements of surface temperature at land sites have been made since at least the sixteenth century, but global coverage has been adequate only since the latter half of the nineteenth century. Routine measurements of sea surface temperature (SST) and marine air temperature have also been carried out since the middle of the nineteenth century,
and the data have appeared in the U.S. Comprehensive Ocean-Atmosphere Data Set (COADS) [Woodruff et al., 1987] and in a compilation made by the British Meteorological Office [Folland et al., 1984; Folland and Kates, 1984; Folland and Parker, 1989]. Several attempts have been made in the past to combine these measurements into hemispheric and global averages, using land measurements alone, ocean measurements alone, or a combination of both [e.g., Mitchell, 1961; Groveson and Landsberg, 1979; Yamamoto and Yoshida, 1980; Folland and Kates, 1984; Folland et al., 1984; Hansen and Lebedeff, 1987, 1988; Jones et al., 1986a, b, c]. These data sets have been discussed critically by Elsaesser et al. [1986], and some of their problems will be mentioned briefly later.

While variations in solar total irradiance can only be detected by sophisticated instruments located above the Earth's atmosphere, sunspots provide an obvious manifestation of solar variability that is easily detected from the ground. Numerous correlations between sunspot number and various weather and climate parameters have been reported in the past and have been reviewed exhaustively in several publications [Herman and Goldberg, 1978; National Academy of Sciences, 1982; McCormac, 1983]. The pitfalls inherent in these correlative studies have been discussed by Pittock [1978], and although several potential mechanisms to explain a relationship between solar magnetic activity and climate have been suggested in a qualitative way, none has so far stood up to quantitative examination. The basic problem is that the energy contained in the extreme ultraviolet, X ray, and energetic particle outputs from the sun that are known to vary dramatically with solar activity represents only a tiny fraction of the total irradiance that drives global climate and, furthermore, even that tiny amount of energy is dissipated in the upper atmosphere and could influence surface climate only indirectly. A discussion of the various suggestions that have been made to avoid these difficulties is beyond the scope of this paper, and the interested reader can consult the references quoted above or more recent papers [e.g., Tinsley et al., 1989].

If the total irradiance were to vary with solar activity, however, a direct relationship with climate would be clear, and this possibility has been raised many times in the past. The most convincing evidence of such a relationship was provided by Eddy [1976, 1977], who pointed out an overall positive correlation between certain indicators of global temperature and the record of solar activity as represented by radiocarbon anomalies over the past millennium, the most striking feature being the coincidence of the Maunder Minimum of solar activity, when sunspots essentially disappeared for about 70 years, and the peak of the so-called Little Ice Age. Eddy [1976] suggested that this relationship could be explained if the solar irradiance had varied in step with the long-term variation of solar activity, i.e., with the envelope of the 11-year sunspot cycle.

This is in essence the possibility that is explored in this paper, using recent data on solar-irradiance variability and the SST record of the past 130 years to show that there is evidence that the irradiance does indeed have a long-term variation with solar activity as represented by the envelope of the 11-year cycle, and that this variation has a large enough amplitude to affect global average SSTs and hence global climate. While solar irradiance variations are not likely to have been the only cause of climate fluctuations, especially in recent decades when the buildup of greenhouse gases in the atmosphere has begun to approach climatically significant levels [e.g., Hansen et al., 1988], estimates of their magnitude is a necessary preliminary to the important task of predicting how soon the global warming caused by man's activities will begin to overwhelm the natural variability of the global climate system.

2. SSTs and Solar Irradiance

In earlier publications [Reid, 1987; Reid and Gage, 1988] we pointed out a fairly striking similarity between the globally averaged SST record of the past 130 years [Folland and Kates, 1984; K. Folland, private communication, 1987] and the long-term record of solar activity, as represented by the 11-year running mean Zürich sunspot number. The two time series are shown in Figure 1 [after Reid, 1987], in which the SST values represent annual departures from an arbitrary zero level, smoothed by the same 11-year running mean filter as the sunspot numbers. While not by any means identical, the two time series have several intriguing features in common. Both show a prominent minimum in the early decades of this century, followed by a steep rise to a maximum in the 1950s, a brief drop during the 1960s and early 1970s, with a final rise. The formal cross-correlation coefficient between the two time series is 0.75, significant at above the 95% level, even after taking the high degree of autocorrelation in the individual series into account [Que- nolle, 1952].

The possibility that the two time series could be related through a long-term variation in the total solar irradiance in phase with solar activity was explored with the aid of a one-dimensional model of the ocean's thermal structure [Hoffert et al., 1980], allowing the solar constant to vary linearly with the running mean sunspot number, and adjusting the amplitude of the variation until the best fit between the observed and modeled SST was achieved. The increase in solar irradiance between 1910 and 1960 required to achieve the fit between the two time series was about 0.6%, giving an average rate of increase of 0.012% per year. Although no reliable measurements of total solar irradiance exist for this period, the magnitude seemed reasonable at the time, since spacecraft measurements of the total irradiance...
were showing a long-period decline of about 0.02% per year from 1980 to 1985 [Willson et al., 1986]. Such a variation, however, would have to be sustained for longer than a decade in order to reach a level at which it would be climatically significant, and the recent continuation of the solar irradiance measurements by both the Active Cavity Radiometer Irradiance Monitor (ACRIM) instrument on the Solar Maximum Mission (SMM) spacecraft and the Earth Radiation Budget (ERB) experiment on Nimbus 7 has shown a levelling off and possibly an increasing trend following sunspot minimum [Willson and Hudson, 1988; Hickey et al., 1988]. These measurements have generally been taken as indicating that the variation in irradiance is directly keyed to the 11-year sunspot cycle, but in fact the most that can be said is that they probably show the existence of a component of irradiance variation that is keyed to the 11-year cycle. Measurements over a much longer time period will be needed to establish the existence or nonexistence of the longer period variations that are likely to be of major importance to climate fluctuations, as has been emphasized by Hoffert et al. [1988]. The existence of a significant 11-year component is predictable, in fact, since about 20% of the total irradiance variation has been estimated to be due to variations in the ultraviolet part of the spectrum [Lean, 1989], which is known to have a substantial 11-year variability.

As mentioned above, measurements of the sun's total irradiance with the precision needed for estimating climatic effects can only be carried out above the atmosphere, since variations in cloud cover and atmospheric transmittance are of much larger magnitude than the irradiance variations. Reliable continuous spacecraft measurements did not begin until 1978, but several spot measurements using similar instruments were made from balloons and rockets at earlier times. These have been summarized by Fröhlich [1987], and Figure 2 shows his compilation of the measurements during the time period 1967–1984 (details of the individual measurements can be found in the work by Fröhlich [1987]). The trend of the individual measurements during the 1980s is very close to that of the continuous spacecraft measurements, lending confidence in their validity. The earlier measurements, however, show much more scatter, but their average is significantly lower than the later values; in fact, eight of the nine early measurements are lower than any of the later measurements. This suggests that either the irradiance really was lower near the peak of solar cycle 20 in 1969–1970 than near the corresponding peak of cycle 21 in 1979–1980, or the earlier measurements were all affected by some common systematic error. Such an error could arise if the residual stratospheric transmittance above the instrument had been overestimated, for example, but this possibility seems unlikely [Fröhlich, 1987]. Evidence supporting a real change in irradiance is provided by the balloon measurements by the University of Denver group [Kosters and Murchay, 1979], who flew two separate radiometer units in the same balloon in 1968, finding an instrumentally related difference of 6 W m⁻² in the measured irradiance between the units. A flight of the same two instruments in 1978 gave irradiance values for both units that were 5 W m⁻² higher than the 1968 values, preserving the instrumental difference.

Another feature of Figure 2 is the difference in the amount of scatter between the two groups of measurements. At least part of the scatter in the earlier measurements can be attributed to the short-term variation in irradiance caused by sunspot blocking, which can cause measurements on individual days near solar maximum to vary by several watts per square meter [Willson and Hudson, 1988]. In fact, the surprising feature of Figure 2 is perhaps the lack of scatter in the later measurements rather than the presence of scatter in the earlier measurements.

Fröhlich [1987] suggested that the observed variation between 1967 and 1985, if real, could be evidence for a 22-year cycle in irradiance with a peak-to-peak amplitude of about 0.4%. This is a valid possibility and may be supported by the report of a significant 22-year periodicity in the global SST time series [Newell et al., 1989]. The observations could just as well be taken as evidence of the existence of a secular variation of solar irradiance with the overall envelope of solar activity [Eddy, 1976; Reid, 1987], however, as suggested by the work described above. Figure 3 shows the envelope of the 11-year sunspot cycle, derived by simply connecting the successive maxima; the curve has been extended to 1990 on the assumption that the maximum of cycle 22 has been passed and that it reached about the same level as the maximum of cycle 21. (Annual values of the envelope have been used in this study as the measure of long-term solar activity, instead of the running mean sunspot number used in the earlier work. The two quantities are closely related.) A plot of individual annual values of the global SST anomalies and the sunspot-cycle envelope is shown in Figure 4. The cross-correlation coefficient between the two time series is 0.71, significant at greater than the 95% confidence level after adjusting the number of degrees of
freedom to allow for autocorrelation in both series [Que- nounille, 1952].

The irradiance measurements in Figure 2 have a mean value of about 1363 W m\(^{-2}\) near the peak of cycle 20 in 1969-1970 and a mean value of about 1367 W m\(^{-2}\) near the peak of cycle 21 in 1979-1980. This increase of 4 W m\(^{-2}\), or 0.3%, took place while the sunspot number envelope increased by 45, from about 105 to 150, as shown by the asterisks in Figure 3. Making the crude assumption that the irradiance is linearly proportional to the envelope, these two values lead to an empirical relationship between irradiance \(S\) and sunspot number envelope \(N\) of the form

\[
S(W m^{-2}) = 1353.6 + 0.089N \tag{1}
\]

The range in \(S\) between the minimum in cycle 14, with \(N = 65\), and the maximum in cycle 19, with \(N = 191\), is 11 W m\(^{-2}\), or about 0.8%. Considering the crude nature of the assumption and all the various uncertainties, this is reasonably close to the range of 0.6% that was needed to achieve the fit between observations and model calculations reported earlier [Reid and Gage, 1988]. The relationship in (1) also predicts that the solar irradiance for zero sunspot number, corresponding to the Maunder Minimum period in the late seventeenth and early eighteenth centuries, should have been about 1354 W m\(^{-2}\), or about 1% below the 1980 value. Eddy [1976] pointed out that this period corresponded to the peak of the Little Ice Age, during which global mean temperatures were substantially lower than today, as evidenced by contemporary climate observations and by such proxy indicators as the advances of mountain glaciers in Europe and North America. The global temperature change between the Little Ice Age and the present was much greater than the change between the Modern Minimum of the early twentieth century and the maximum of the 1950s, yet the differences in the hypothesized solar irradiances (1% versus 0.8%) were comparable. The reason for the difference probably lies in the long duration of the Maunder Minimum coupled with the thermal inertia of the oceans, as we shall discuss in the next section.

3. Model Calculations

The relationship (1) between solar irradiance and the envelope of the sunspot number as shown in Figure 3 has been used to calculate the globally averaged SST over the past three centuries using a one-dimensional model of the thermal structure of the global ocean. As in the earlier work [Reid, 1987; Reid and Gage, 1988], the model used was that developed by Hoffert et al. [1980] in which a mixed layer of 100-m depth is coupled to a deep ocean of 4000-m depth. The mixed layer is heated directly by solar radiation and cooled by radiation to space, by transfer of heat to the atmosphere, and by eddy transport of heat to the deep ocean, while a thermohaline circulation driven by processes occurring in a polar sea provides upwelling of cold water from the base of the deep ocean. Details of the model can be found in the work by Hoffert et al. [1980], and only a brief outline will be given here.

In a steady state the longwave radiation to space, \(F\), must be balanced by the absorbed incoming solar radiation, i.e.,

\[
F = \frac{S}{4}(1 - \alpha) \tag{2}
\]

where \(S\) is the solar irradiance and \(\alpha\) is the globally averaged albedo of the Earth.

Assuming a linear relationship between surface temperature and outgoing radiation to space [Budyko, 1969], and adopting the logarithmic relationship between surface temperature and atmospheric carbon dioxide concentrations suggested by the radiative-convective model of Augustsson and Ramanathan [1977], Hoffert et al. [1980] express \(F\) as

\[
F(T_s, c) = A + B(T_s - T_0) - C \ln (c/c_0) \tag{3}
\]

where \(T_s\) and \(c\) are surface temperature and \(\text{CO}_2\) concentration, respectively, \(T_0\) and \(c_0\) are baseline reference values, and \(A\) is the longwave radiation corresponding to the baseline state. The coefficient \(B\) expresses the sensitivity of \(F\) to surface temperature changes, and the value obtained from the model calculations of Augustsson and Ramanathan [1977] depends on the assumed water vapor distribution and cloud top temperature. An average value of 2.2 W m\(^{-2}\) K\(^{-1}\) was adopted, following Hoffert et al. [1980]. The value of \(C\), which has to be determined from the results of model studies of the impact of changing greenhouse gas concentrations on surface temperature, will be discussed later.

Inserting the equilibrium value of \(F\) from (2), the equilibrium surface air temperature \(T_{eq}\) can be expressed as

\[
BT_{eq} = BT_0 + (S/4)(1 - \alpha) - A + C \ln (c/c_0) \tag{4}
\]

\(T_{eq}\) is the surface air temperature that would ultimately be reached if the incoming radiation remained unchanged for a long period of time. Because of the thermal inertia of the ocean, however, the actual surface air temperature \(T_s\) in a period of changing radiative flux lags behind the equilibrium temperature, and also differs from the ocean mixed layer temperature \(T_m\). For the purposes of the model, this difference is assumed to be constant, so that \(dT_s/dt = dT_m/dt\). It can then be shown [Hoffert et al., 1980] that the time variation of \(T_s\) can be expressed as

\[
dT_s/dt = (T_{eq} - T_s)/\tau - (\partial Q/\partial t)/h_m \tag{5}
\]

where \(\tau(B, h_m)\) is the time constant for radiative equilibrium of the ocean mixed layer, \(h_m\) is the depth of the mixed layer, and \(Q\) is proportional to the integrated heat content of the deep ocean, i.e.,
where \( T_d \) is the deep-ocean temperature, \( z \) is the depth below the base of the mixed layer, and the integration extends from \( z = 0 \) to \( z = h_d \), the bottom of the ocean. A change in the solar irradiance changes \( T_{\text{surf}} \), by virtue of (4), and hence \( d T_{\text{surf}} / d t \). Integration of (5) then determines \( T_p \) as a function of time and can be carried out once the deep-ocean term is evaluated.

The temperature changes in the deep ocean are governed by diffusion of heat from the mixed layer and advection of heat (or cold) from the ocean bottom, driven by the thermohaline circulation. The appropriate equation is

\[
\frac{\partial T_p}{\partial t} = K \frac{\partial^2 T_p}{\partial z^2} + w \frac{\partial T_p}{\partial z}
\]

(7)

where \( K \) is an eddy diffusion coefficient and \( w \) is the globally averaged thermohaline upwelling velocity. Hoffert et al. [1980] estimated these parameters on the basis of empirical measurements of tracer transport in the ocean and rates of formation of ocean deep water. Their values, which have been adopted here, are \( K = 2000 \text{ m}^2 \text{ yr}^{-1} \) (\( = 0.6 \text{ cm}^2 \text{ s}^{-1} \)) and \( w = 4 \text{ m yr}^{-1} \). Equation (7) is subject to boundary conditions at the base of the mixed layer (\( z = 0 \)), where \( T_d = T_m \), and at the bottom of the ocean, where the heat flux from advection and diffusion must balance the flux from the polar sea, whose temperature is given by \( T_p \). This lower boundary condition is expressed as

\[
K \frac{\partial T_p}{\partial z} + w T_p = w T_p
\]

(8)

at \( z = h_d \). The polar sea temperature \( T_p \) is held fixed at a value of 1.2°C.

Using the time-varying solar "constant" given by (1), the coupled differential equations (5) and (7) have been solved numerically subject to the boundary conditions given above. The globally averaged albedo \( \alpha \) in (4) was taken as 0.3, the equilibrium time constant \( \tau \) as 3.9 years, and the constant difference between surface air temperature and mixed layer temperature as 4.8°C, following Hoffert et al. [1980]. The baseline state chosen corresponded to modern conditions, that followed, and the strong warming that is continuing into the present. The temperatures of the recent decades are in general agreement with observations of surface air temperatures [Hansen and Lebedeff, 1988; Jones et al., 1988]. As mentioned in the introduction, the 1850s mark the start of reasonably reliable temperature records on a wide enough scale that hemispheric or global averages can be at least estimated. Prior to that time, only rough guesses can be made on the basis of such indirect evidence as the extent of mountain glaciers or the length and severity of mid-latitude winters. There can be little doubt of the existence of the Little Ice Age, but whether or not its minimum temperatures occurred in approximate coincidence with the Maunder Minimum period is about 1°C cooler than today, in general agreement with other estimates [e.g., Wigley, 1988].

The dip in temperature in the early nineteenth century studied in detail by Harvey and Schneider [1985]. The tests carried out on our version of the model are essentially in agreement with their results and will not be repeated here. The most interesting parameter from this point of view is the upwelling speed \( w \), which affects the mixed-layer temperature in a rather complex way. Increases in \( w \) tend to increase the amplitude of the response to a changing solar irradiance, giving a more rapid warming when the irradiance is increasing, and a more rapid cooling when the irradiance is decreasing. For reasonable changes of \( w \), however, the effect is relatively small. The possibility that \( w \) may be coupled to the mixed-layer temperature, however, leads to interesting feedback possibilities that have been explored by Harvey and Schneider [1985], Watts [1985], and Gaffin et al. [1986].

The calculated SST variation for the period 1660–1990 is shown by the continuous curve in Figure 5. (The version of the SST record used here is that current in 1987 [C. K. Folland, private communication, 1987]. A more recent version [Folland and Parker, 1989] has a somewhat reduced amplitude but preserves the essential features of the variation.) The time series of observed SST anomalies from 1856 through 1986 has been minimally smoothed with a simple 1-2-1 filter, and plotted on top of the calculated curve to illustrate the overall agreement. The model curve is seen to reproduce several of the observed features, including the cooling that occurred around the turn of the century, the strong warming of the 1920–1960 period, the brief cooling that followed, and the strong warming that is continuing into the present. The temperatures of the recent decades are calculated to be the warmest since the seventeenth century, in general agreement with observations of surface air temperatures [Hansen and Lebedeff, 1988; Jones et al., 1988].
shown in Figure 5 is a result of the so-called Dalton Minimum of solar activity that occurred at that time, and is an interesting case. The famous “year without a summer” of 1816 occurred during this period and has often been attributed to the massive Tambora volcanic eruption in 1815 [e.g., Lamb, 1970; Landsberg and Albert, 1974], but Angell and Korshover [1985] found only weak evidence for a large-scale cooling following the Tambora eruption. Schneider [1983] has pointed out that abnormal meteorological conditions, such as an unusual southward dip of the summertime jet stream, were probably needed in order to explain the occurrence of freezing temperatures in July over a large area of the northern middle latitudes. It is not clear how the aerosol cloud from a tropical volcano could produce such an effect.

The northern hemisphere temperature estimate of Grovenan and Landsberg [1979] actually shows a steady drop in temperature beginning about 1780 and reaching its minimum level about 1815, the year of the Tambora eruption. It would be unfair, however, to quote this favorable feature of the Grovenan and Landsberg [1979] record without also quoting the fact that the Maunder Minimum does not appear as a particularly cold period, as pointed out by Robock [1979].

All attempts to reconstruct hemispheric or global temperature variations beyond the most recent century and a half are necessarily highly uncertain and any conclusions drawn from them must be treated with great caution.

4. Terrestrial Sources of Climate Variability

Considering the crudeness of the solar irradiance model and of the one-dimensional ocean thermal model, and the major uncertainties associated with averaging the SST observations over the globe, the agreement shown in Figure 5 might be regarded as largely fortuitous, especially since other internal and external sources of global climate variation have been ignored. Chief among the external sources are the enhanced greenhouse effect from CO2 and other radiatively active gases in the atmosphere and the reduction of solar radiation at the surface by volcanic aerosol clouds in the stratosphere.

The volcanic contribution to climate variability on the decadal time scale has been debated for years and remains controversial. Significant short-term reductions in surface temperature following individual volcanic eruptions in recent decades have been reported by several authors [e.g., Newell, 1984; Taylor et al., 1980; Rampino and Self, 1982; Sear et al., 1987], but Angell and Korshover [1985] found that the evidence for these reductions was weak for all six major eruptions between 1780 and 1980. Robock [1978], on the other hand, found good agreement between the global temperature record compiled by Mitchell [1961] and a model calculation using published dust-veil indices [e.g., Lamb, 1970]. A critical factor in the assessment of volcanic effects is the residence time of the aerosols in the stratosphere, which is probably only 1 or 2 years. The thermal inertia of the ocean would severely damp the SST response to forcing on such a short time scale, providing some justification for ignoring the volcanic contribution. Since land surface temperatures would be expected to respond more readily, however, the contribution of volcanic aerosols may help to account for the discrepancies between land-based surface air temperature and SST records.

More critical is the neglect of the greenhouse-gas contribution. Assuming that this has all been due to carbon dioxide, ground-based measurements of CO2 concentration can be used to derive the time variation of c in equation (4), and hence to determine the variation in equilibrium temperature \( T_{eq} \) caused by the increasing CO2 burden. The constant C can be derived from the results of model simulations employing a full radiative treatment. Many such calculations have been made, and the parameter commonly used to compare them is the predicted equilibrium temperature change caused by doubling the CO2 concentration. Schlesinger and Mitchell [1987] have reviewed the various model treatments, and their tabulation shows that the predicted \( \Delta T_{eq} \) for doubled CO2 ranges all the way from values less than 1°C to values greater than 4°C, depending on the assumptions made in the model. While most models appeared to be agreeing on the higher values until recently, treatments that take account of the radiative effects of phase changes in cloud water have substantially reduced the predicted value of \( \Delta T_{eq} \) to less than 2°C [Mitchell et al., 1989].

Hoffert et al. [1980] used a simple exponential fit to the observed CO2 concentrations of the form

\[
c(t) = 290 + 26 \exp [(t - 1958)/34] \tag{9}
\]

where \( c(t) \) is the CO2 concentration in ppmv and \( t \) is the year. This relationship seems to have continued to give a good fit to more recent data extending through 1987 [e.g., Mitchell, 1989] and has been used in equation (4) together with the baseline value of \( c_0 = 340 \) ppmv.

Figure 6 shows the SST variation calculated from the model for values of \( \Delta T_{eq} \) for doubled CO2 of 1°C, 2°C, and 4°C, corresponding to values of C (from (4)) of 3.18, 6.35, and 12.7 W m\(^{-2}\), respectively. The solar irradiance was kept constant and small adjustments were made in its magnitude for each curve in order to produce a temperature of 19.6°C in 1980, close to the observed value. Since the model adopted for the CO2 variation has a simple exponential growth factor, all of the curves show a monotonically increasing temperature, and it has been suggested that the observed rise in temperature from 1920 through about 1960 may be a signal of the greenhouse warming. The greenhouse effect cannot, however, explain the subsequent cooling nor the behavior of the temperature record in the late nineteenth century. The three curves in Figure 6 show temperature increases of 0.07°C, 0.14°C, and 0.26°C, respectively, between 1920 and 1960, while the observed temperature increase was about 0.5°C. The enhanced greenhouse effect should therefore have been
a significant contributor to the mid-twentieth century warming but was probably not the dominant one. That situation is likely to change within the next few decades, however, if current model predictions are correct [e.g., Hansen et al., 1988], since the three curves in Figure 6 predict temperature increases from 1980 to 2000 of 0.13°, 0.25°, and 0.49°C, respectively.

Apart from the possible external sources of climate change, internally generated changes can take place in such a highly nonlinear dynamical system as that of global climate. Internally generated temperature variations of the same magnitude as that of the recent climate record have been produced in several models [e.g., Hasselmann, 1976; Robock, 1978; Hansen et al., 1988], and the possibility that the temperature record shown in Figure 1 represents nothing more than random stochastic variations occurring in a complex system remains open. Globally averaged quantities, however, may be less subject to such random variations than short-term or localized parameters, since they tend to be constrained by rather simple global relations, such as the balance between received solar radiation and emitted long-wave radiation, for example.

5. The Temperature Record

The globally averaged SST has been used in this study as a proxy for the globally averaged planetary surface temperature, and the question of the validity of this assumption deserves some discussion. The related question of whether SSTs or surface air temperatures are the best proxies has been debated at length and remains controversial. All else being equal, there is little doubt that SSTs would be the better choice, since oceans cover more than 70% of the planet’s surface, and since SSTs have a larger spatial coherence and less temporal variability than land temperatures. The most extensive surface air temperature compilations to date are those of Jones et al. [1986a, b] and Hansen and Lebedeff [1987, 1988], and they show good overall agreement with each other (which is not unexpected, since there is a considerable amount of overlap in the individual station temperature records used). Both of these temperature records are significantly different from the SST and marine air temperature records that form the British Meteorological Office compilation [Folland et al., 1984] and the U.S. Comprehensive Ocean-Atmosphere Data Set [Woodruff et al., 1987; Oort et al., 1987], which also agree well with each other but which also overlap to a considerable extent in their data sources. Jones et al. [1986c] attempted to resolve the discrepancies by adjusting marine air temperature records from ocean locations near coasts to agree with adjacent coastal land air temperatures wherever suitable paired records were available, and then adjusting the SSTs to agree with the marine air temperature. This procedure is somewhat arbitrary and effectively assumes that all of the error lies in the measurement of SST and marine air temperature. There are in fact reasons to suspect both data sets to some extent: the ship-based measurements because of changes in measurement technique, ship speed, and height above the ocean surface, and the land-based measurements because of changing locations, local times of observation, and urban growth. While much effort has been made to minimize these effects, a considerable amount of uncertainty must remain in both data sets.

To illustrate the problem, Figure 7 shows the result of filtering the SST time series used in this paper and the surface air temperature (SAT) time series of Hansen and Lebedeff [1987, 1988], using a simple Gaussian low-pass filter. Both series are plotted as departures from an arbitrary zero level.

Fig. 7. Low-pass filtered series of globally averaged surface air temperature [Hansen and Lebedeff, 1987, 1988] and of globally averaged SST [Folland and Kates, 1984]. Both are plotted as departures from an arbitrary zero level.

Clearly, the reliability of the temperature record is a crucial factor in any attempt to identify the major sources of climate forcing, and different temperature records can lead one to entirely different conclusions. Robock [1978], for example, estimated the relative contribution of solar constant variations, volcanic eruptions, and anthropogenic effects on the temperature record as compiled by Mitchell [1961], using a model solar irradiance variation very similar to that of equation (1). He concluded that solar variability was not important and that the largest fraction of the variance could be explained by volcanic eruptions. The principal reason for this discrepancy appears to be the fact that the temperature record of Mitchell [1961], which was based on land measurements, did not show the early twentieth century cooling, which is a pronounced feature of the SST record. The reality of that cooling is crucial to the
hypothetical presented here, but it is difficult to establish unequivocally. The fact that it is present at all latitudes between 40°N and 40°S [Oort et al., 1987] is strong evidence that it represents a real global cooling and is not an artifact of the averaging process, of changes in global shipping lanes, or of changes in the conditions or techniques of measurement. As far as the validity of the global averaging is concerned, Oort et al. [1987] have shown with the COADS data set that the global coverage of SST measurements has not changed greatly since the late nineteenth century, at least in the latitude range 40°N to 40°S. The principal change has been in the number of measurements available rather than in the actual locations in which the measurements have been made, although there has been a significant growth in the quantity of high-latitude data in recent decades.

6. SOLAR LUMINOSITY VARIATIONS

It is natural to ask whether irradiance variations of the magnitude suggested here (≈0.1 to 1% on time scales of decades to centuries) are permissible from the viewpoint of contemporary solar physics. Unfortunately, the answer is not clear. The radiation escaping from the solar photosphere is largely determined by convective heat transport occurring in the sun’s outer layers, about whose properties little direct evidence exists. Models of the convective zone generally use a mixing-length approach to relate the heat flux to the temperature gradient, and simulated perturbations in the mixing length have been shown to have a significant impact on the luminosity [Dearborn and Newman, 1978; Gilliland, 1982; Endal et al., 1985], particularly if they occur in the outermost superadiabatic layer of about 3000-km thickness. Such perturbations might result from the presence of time-varying magnetic fields in the outer convective zone. Since the strength of these magnetic fields is likely to be directly related to the size of the approaching solar activity peak, a direct connection between the sunspot-cycle envelope and luminosity would not be surprising. Luminosity variations of 0.1-1% can be simulated by quite reasonable perturbations in convective transport [Dearborn and Newman, 1978] and would not necessarily be accompanied by variations in the radius of the sun that could be detected, for example, by eclipse observations. Variations in solar radius on the time scale of decades to centuries have been reported [Gilliland, 1981; Sofia et al., 1983], however, and are likely to be accompanied by luminosity variations of undetermined size (or even sign).

In a study of the solar irradiance variations seen since 1980 by spacecraft radiometers, Foukal and Lean [1988] and Lean and Foukal [1988] have found that the residual variations after accounting for the sunspot blocking effect are positively correlated with indices of facular intensity, suggesting that the darkening of the sun produced by sunspots is compensated by brightening of the associated faculae and of the background network of facular elements. The long-term positive correlation of total irradiance with solar activity then implies that the facular brightening exceeds the sunspot darkening. More recently, Foukal and Lean [1990] have constructed a model of total irradiance variation extending back to 1874, assuming that all of the variation has been due to the causes identified during the decade of direct spacecraft measurements, i.e., sunspot blocking and facular enhancement, and using sunspot number and sunspot area as separate proxies for these two variables. Since their model does not allow for any total irradiance variation other than that directly associated with sunspot number and area, both of which are dominated by the 11-year cycle, their empirical irradiance variation is not relevant to the long-term variation suggested here. The fact that the spacecraft measurements since 1980 have not shown any evidence for a long-term variation other than that due to the 11-year cycle itself may simply be a result of the near constancy of the envelope during recent years, as shown in Figure 3. The long-term variation in solar activity appears to be at a turning point.

There is general agreement that irradiance variations of the order of 0.1% occurring over an 11-year cycle are very unlikely to have significant climatic effects [Hoffert et al., 1988], largely because of the severe damping of such short-term variations by the thermal inertia of the oceans. This is illustrated in Figure 8, which shows the attenuation and lag in mixed-layer temperature calculated from the model for irradiance variation as functions of the forcing period. The irradiance was allowed to vary sinusoidally with an amplitude of 1%, and the model was run for long enough to settle down to a repeatable sinusoidal response. The attenuation is defined here as the ratio of the amplitudes of the resultant mixed-layer temperature and equilibrium temperature cycles, and the lag is the delay time between peaks in the two cycles. Clearly a century-scale variation, with an attenuation factor of about 0.75, has a major advantage over both 11-year and 22-year cycles, with attenuation factors of 0.25 and 0.43, respectively.

Our understanding of the processes that occur in the sun’s convective zone is still in a fairly rudimentary state, and it is not present possible to set realistic limits on the magnitude of long-term variations in solar luminosity on theoretical grounds alone. Rapid progress is being made in both theoretical and observational solar physics, however, and it seems not unreasonable to expect some quantitative estimates within a relatively short time.

7. SUMMARY AND CONCLUSIONS

The record of globally averaged sea-surface temperatures over the past 130 years has been shown to have a closely
similar variation to that of solar activity as derived from the envelope of the sunspot cycle. This finding essentially updates the suggestion by Eddy [1977] that long-term changes in solar activity over the past several centuries have been accompanied by changes in global climate. The simplest explanation of this correspondence lies in a variation of the sun’s total irradiance (the solar “constant”) with solar activity as proposed by Eddy [1977], and calculations with a one-dimensional model of the oceanic thermal structure have shown that the required variations in solar irradiance are of the order of 0.1–1%. Solar irradiance measurements made near the peak of the last two sunspot cycles seem to show that variations of the required magnitude may indeed have taken place.

It must be admitted, however, that the evidence supporting the hypothesis is largely circumstantial, and the supporting arguments are less defensible than one could wish. As mentioned earlier, the SST record itself has serious inherent uncertainties associated with it, and one could quarrel with the meaning of the global average SST as a measure of global climate, and with the applicability of a simple one-dimensional model of the thermal structure of the global ocean. Even taking these caveats into consideration, however, the evidence does seem to point toward solar irradiance variations having played a major role in climate variations over the past century and a half, although it seems unlikely that they were the sole cause. As far as the future is concerned, there is no obvious reason to question the predictions of future global warming due to the atmosphere’s increasing burden of greenhouse gases. Unless there are serious deficiences in our understanding, the conclusion that the greenhouse warming will eventually overwhelm other natural effects of climate variability seems inescapable. The critically important question of when this will happen remains unanswered, and its answer depends on how accurately we can estimate the natural internal and external causes of climate variability.

REFERENCES


Lean, J., Contribution of ultraviolet irradiance variations to changes in the sun’s total irradiance, Science, 244, 197-200, 1989.


—G. C. Reid, Aeronomy Laboratory, National Oceanic and Atmospheric Administration, 325 Broadway, Boulder, CO 80303.

(Received February 7, 1990; revised August 24, 1990; accepted October 10, 1990.)