PERGAMON

Journal of Atmospheric and Solar-Terrestrial Physics 61 (1999) 3–14

Journal of
ATMOSPHERIC AND
SOLAR-TERRESTRIAL
PHYSICS

Solar variability and its implications for the human environment

George C. Reid*

Aeronomy Laboratory, National Oceanic & Atmospheric Administration and Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, CO 80303, U.S.A.

Received 29 October 1997; accepted 31 July 1998

Abstract

Solar variability can affect human activities in a variety of ways, from changing our climate to disrupting power distribution facilities and shortening the orbital lifetime of satellites. This tutorial paper will be concerned only with effects on the surface environment that can have a direct impact on our everyday life, such as variations in the stratospheric ozone layer that shields us from harmful ultraviolet radiation, and changes in global climate that can hinder or delay the detection of climate changes that might result from our own technological activities. The emphasis is on potential mechanisms, rather than on reported correlations between solar and terrestrial parameters, but reference to certain observations will be made. Realization of a potential impact of solar variability on our local environment has progressed a long way in the last few decades, from denial to partial acceptance, but a complete assessment of its reality and magnitude remains a distant goal.

1. Introduction

The paramount importance of the Sun to human existence is clear to everyone, yet comparatively little attention has been paid to the influence of solar variability on the human environment until quite recent times. Part of the recent upsurge of interest is due to the fact that variations in the portion of the Sun's output that reaches the Earth's surface have now been seen, and part to the realization that human activities may be starting to have a detectable impact on our climate, with the prospect of an ever-increasing impact for the future. Detection and prediction of these anthropogenic effects require a knowledge of the natural background of variability that can disguise their signal, and solar variability is one potentially important contributor to this natural background. The purpose of this paper is not to review the many claims of correlations between solar activity and climatic variables that have appeared in the literature over the years, but rather to discuss the principal mechanisms that have been suggested as contributing to a relationship between solar variability and the Earth's surface environment. Since the discussion is intended to be tutorial in nature, only a few key references to the literature will be given, and the interested reader can consult these for more detailed discussion and for links to the more extensive literature.

The Sun's output varies on an enormous range of time scales, from minutes in the case of flares and other manifestations of surface magnetic activity, to the billion-year time scale of solar evolution. While connections between solar variability on the shorter time scales and day-to-day weather changes have been suggested, the main interest in terms of a comparison with anthropogenic effects is on the time scale of decades and longer, and our discussion here will be concerned with changes on those time scales, including in particular the 11-year solar activity cycle.

The largest variations take place in the extreme ultraviolet and X-ray regions of the solar spectrum and in the energetic charged particle emissions, all of which originate well above the photosphere. Their effects are mainly felt in the Earth's thermosphere and ionosphere, where they have an impact on satellite technology and less directly on facilities such as electrical power distribution lines. These 'space weather' effects will not be discussed here, and our emphasis will be on the more subtle aspects

* Tel.: +1-303-497-3304; fax: +1-303-497-5373.
E-mail address: reid@al.noaa.gov. (G.C. Reid)

S1364–6826/99/$ - see front matter
PII: S1364–6826(98)00111–4

of solar variability that have a direct impact on our immediate environment at the surface of the planet.

Three major categories of solar variability are involved:

(1) variations of the spectral irradiance, especially in the near and middle ultraviolet, leading to changes in the ultraviolet environment at the Earth's surface, and possibly also to variations in tropospheric dynamics;
(2) variations of the Sun's total irradiance (the solar 'constant'), leading to changes in the planetary radiation budget, and to variations in regional and global climate;
(3) variations in the solar wind, leading to changes in cosmic-ray ionization and the global electric circuit, with potential consequences for cloud nucleation and growth.

The current status of our understanding of each of these aspects of solar variability, and their impact on the surface environment, will be summarized in the sections below.

2. Ultraviolet spectral irradiance variability

Except for the technological impacts mentioned above, events taking place in the mesosphere or thermosphere have little, if any, direct influence on the human environment. Accordingly, only the effects of solar variability on the stratosphere and troposphere will be considered, i.e., our attention will be confined to altitudes in the atmosphere below the stratopause, at about 50 km altitude, or the 1-hPa pressure level. The corresponding region of the solar spectrum consists of wavelengths longer than about 175 nm, since shorter wavelengths are mostly absorbed above this level. Solar radiation in the band between 175 and 242 nm is mainly responsible for the production of stratospheric ozone, which protects the surface of the Earth from the fluxes of ultraviolet radiation at longer wavelengths that would otherwise make life impossible. A few numbers here are instructive. The 175–242 nm wavelength band comprises about 0.13% of the total solar irradiance, or about 1.8 W m^{-2}, while the longer wavelength band extending to 300 nm that is harmful to living organisms comprises nearly 1% of the total, or about 12 W m^{-2}. The atmosphere has thus developed an extremely efficient means of protecting us from ultraviolet radiation: by using 1.8 W m^{-2} of radiation to produce ozone, it shelters us from 12 W m^{-2} of harmful radiation, thanks to ozone's large cross section for dissociation in this wavelength region. Furthermore, the quantity of ozone involved is remarkably small, amounting to a layer only about 3 mm in thickness at a pressure of 1 atmosphere, yet it is responsible for heating the global stratosphere and providing the major driving force for much of the stratospheric wind system.

Since solar radiation is responsible both for creating stratospheric ozone and for producing the radiation from which ozone protects us, the question of long-term variability in the respective irradiance bands is an important one that has attracted a considerable amount of attention in recent years. Figure 1, adapted from Lean (1991) with updates based on recent satellite data (G.J. Rottman, private communication), shows a rough estimate of spectral irradiance variability associated with the 11-year sunspot cycle. The stratospheric ozone production region of the spectrum varies at the 3–10% level, which is considerably more than the variability in the region mainly responsible for ozone destruction, which is at the level of only 1–3%. Little is known about spectral irradiance variability in the visible and infrared regions, shown by the broken line in Fig. 1, but the variability in the total irradiance of the order of 0.1% over the 11-year cycle, discussed in the next section, indicates that spectral irradiance variability in the wavelength band that carries the bulk of the Sun's total output must be of the same order of magnitude. In terms of energy flux, however, a variation in the visible and infrared of 0.1% amounts to about 1.4 W m^{-2}, and is larger by an order of magnitude than the variation in energy flux associated with the much more variable ultraviolet spectral regions. The question of variability in the visible and infrared spectral regions deserves more attention, since the climate system is likely to be quite sensitive to the relative distribution of energy across the spectrum. The Earth's albedo is wavelength dependent, for example, and the absorption of radiation by the oceans is extremely sensitive to wavelength. Blue radiation penetrates to depths of 100 m or more, while radiation in the near infrared is absorbed in the surface skin. Sea-surface temperatures are thus likely to respond differently to changes of the same magnitude taking place in the blue or red ends of the spectrum. Spacecraft programs now in operation or planned for the future should provide us with more information on this question.

The relationship between solar activity and stratospheric ozone has been the subject of many investigations over the years, based on both model simulations and observations. While there is general agreement that the solar-cycle effect on the total ozone vertical column is small—of the order of 1.5 to 2%—the models and observations lead to different conclusions regarding the variation of ozone with height. The situation has been summarized recently by Hood (1997), who used satellite data to show that the bulk of the solar-cycle variability occurred in the lower stratosphere below 30 km, where ozone is controlled by dynamical effects, with only a very small contribution from the middle stratosphere between 30 and 40 km, where models predict a substantial effect (e.g., Brasseur, 1993; Haigh, 1994). While it is tempting to take this discrepancy as evidence of a response of the

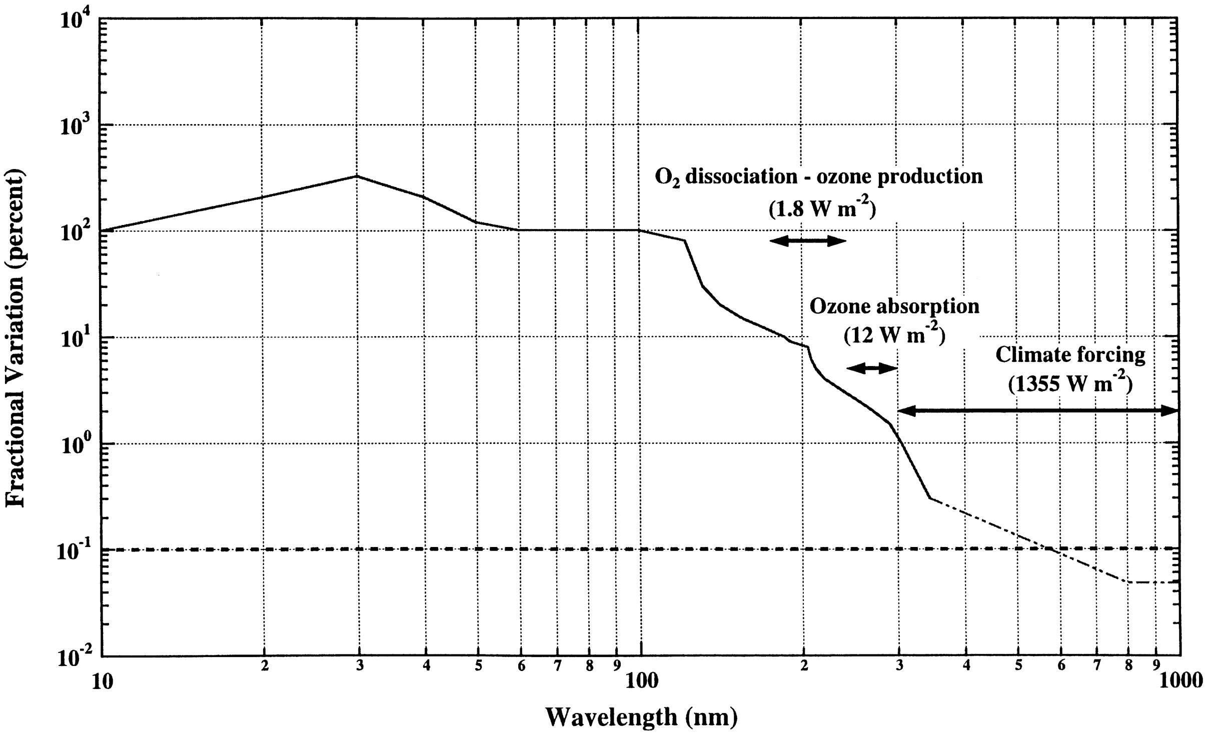

Fig. 1. Fractional variation in solar radiation associated with the 11-year solar-activity cycle.

dynamics of the lower stratosphere to the 11-year solar cycle, both the models and the observations need further refinement before such a conclusion can be fully supported. The possible importance of dynamics has also been shown by Labitzke and van Loon (1997), who used satellite data to show a highly significant correlation between total ozone and the sunspot cycle. The maximum correlation occurred in the subtropics, and not in either the equatorial regions, where ozone production is greatest, or at higher midlatitudes, where its concentration is greatest. The region of maximum correlation was found in the summer hemisphere, shifting with the season, and the highest correlation over the equatorial region was found in the northern summer. The authors suggest that the solar-cycle effect arises through a variation in the meridional transport of ozone, rather than in its chemical production or loss. Meridional transport of ozone by the Brewer-Dobson circulation (Brewer, 1949), however, maximizes in the winter hemisphere, where planetary wave activity is greatest, and on a global scale the peak transport occurs during the northern winter (e.g., Rosenlof, 1995). The fact that the maximum correlation occurs in the summer hemisphere, where transport is weakest, might be taken as arguing against a strong solar influence on transport. Clearly the situation needs further study.

Some general circulation model calculations have shown significant tropospheric effects resulting from variations in solar ultraviolet flux at the ozone-generating wavelengths (e.g., Kodera et al., 1991; Rind and Balachandran, 1995), but the flux variations used have generally been much larger than those observed. Haigh (1996), however, found that significant changes in the dynamics of the lower stratosphere and troposphere resulted from realistic solar-cycle variations of ultraviolet spectral irradiance and ozone in a GCM simulation for perpetual January conditions. During solar maximum conditions, warming of the lower stratosphere gave rise to stronger easterly winds in the summer hemisphere, and penetration of these winds into the upper troposphere caused changes in the tropical Hadley circulation and the midlatitude storm tracks. These results are in general agreement with some of the correlations that have been reported between solar activity and atmospheric properties (Brown and John, 1979; Labitzke and van Loon, 1992), but the magnitude of the model effects is smaller than those of the observations.

In terms of the influence of solar variability on the surface environment, which is the basic theme of this paper, the small 1.5–2% variations in the total ozone column are likely to have a fairly negligible effect on surface ultraviolet fluxes. While present levels of anthropogenic chemical destruction are thought to cause background ozone depletions that are comparable with these solar-cycle variations over midlatitude and tropical locations, much larger anthropogenic effects occur in the polar springtime ozone holes and globally as a result of

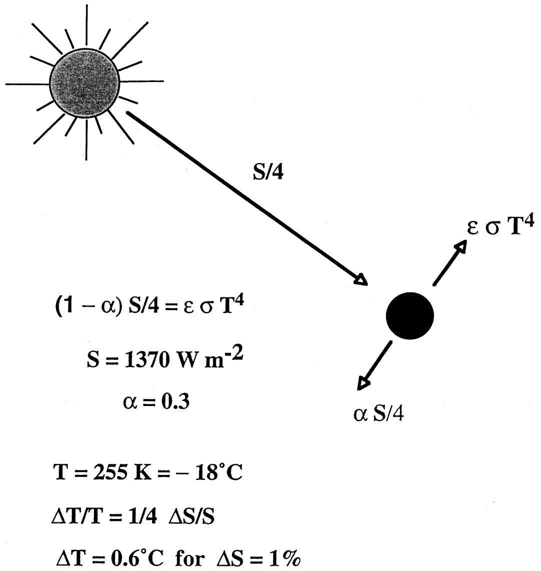

Fig. 2. A simplified sketch of the Earth's radiation balance.

major volcanic eruptions (Hofmann and Solomon, 1989). The possible influences of solar-related changes in stratospheric ozone on tropospheric dynamics are much more speculative, but are potentially more important than the ultraviolet variations from the point of view of the human environment. More effort is needed to place these proposed connections on a more secure footing.

3. Variations in total irradiance

High-precision measurements of the Sun's total irradiance have been carried out routinely from spacecraft since the early 1980s (e.g., Willson and Hudson, 1991). In addition to variations on the time scale of days associated with the appearance of magnetically active regions and their rotation across the solar disk, a variation at the level of about 0.1% has been found, in phase with the 11-year sunspot cycle. Variations in the UV and shorter wavelengths that are absorbed above the troposphere account for about 20% of this variation (Lean, 1989), but the remaining 80% is in the climatically important region of the spectrum.

Figure 2 shows a simple picture of the Earth's radiation balance, in which S is the Sun's total irradiance, $S/4$ is the average irradiance at the top of the atmosphere (the factor 4 being the ratio of the surface area to the cross section), α is the albedo, and ε and σ are the effective emissivity (usually taken as approximately 1) and the Stefan–Boltzmann constant, respectively. T is the effective radiating temperature, which in the case of the Earth is about -18°C, or roughly the actual temperature in the troposphere at an average height of about 6 km. In simple terms, one can think of this as the level below which the atmosphere is too optically thick at this blackbody temperature to be able to radiate to space, due to the presence of such major infrared absorbers as water vapor and carbon dioxide. The surface temperature must adjust to the temperature necessary to maintain a temperature of -18°C at the effective radiating level, given the existing lapse rate, which is determined by processes acting entirely within the atmosphere. If the Sun's output changes, the effective radiating temperature must also change at a rate of about 0.6°C for a 1% irradiance change (although the adjustment can be slow due to the thermal inertia of the oceans), and the surface temperature must follow along assuming that the lapse rate is unchanged. The solar-cycle variation of about 0.08% in the total irradiance reaching the lower atmosphere would then give rise to an equilibrium temperature change of only about 0.05°C. Of course, other factors than variations in solar irradiance can give rise to changes in the equilibrium energy budget, and the addition of so-called greenhouse gases to the lower atmosphere is a particular example. If the lower atmosphere becomes more optically thick in the infrared as a result of such additions, the effective radiating level in the atmosphere must rise in altitude, but the effective radiating temperature remains the same if the solar irradiance is unchanged. The negative lapse rate in the troposphere then forces the temperature to increase at all levels, giving rise to so-called greenhouse warming.

These radiative balance arguments refer to equilibrium conditions at the top of the atmosphere, but the balance at the Earth's surface is considerably more complicated. Figure 3 shows the result of a recent estimate of the energy budget based on the best available data from satellites and other sources (Kiehl and Trenberth, 1997). The net incoming solar radiation at the top of the atmosphere of 342 W m^{-2} translates to 168 W m^{-2} absorbed by the surface, after albedo and atmospheric absorption have been taken into account. The infrared radiation from the surface of 390 W m^{-2} is offset by 324 W m^{-2} of back radiation from the atmosphere and clouds, leaving a net infrared emission of only 66 W m^{-2}. The imbalance in the surface heat budget between 168 W m^{-2} of incoming flux and 66 W m^{-2} of outgoing flux is accounted for by 102 W m^{-2} of cooling by convection and evaporation, largely from the tropical oceans. This does not represent a loss of energy to the system, but merely a transfer of heat from the surface to the atmosphere, since the energy loss from the surface by evaporation is recovered by the release of latent heat in forming clouds. The efficient transfer of heat from the surface to the atmosphere also helps to maintain the temperature lapse rate close to the moist or dry adiabatic value, which is much less steep than the purely radiative lapse rate. As a result, the greenhouse warming at the surface is considerably less than it would be in an atmosphere in radiative equilibrium.

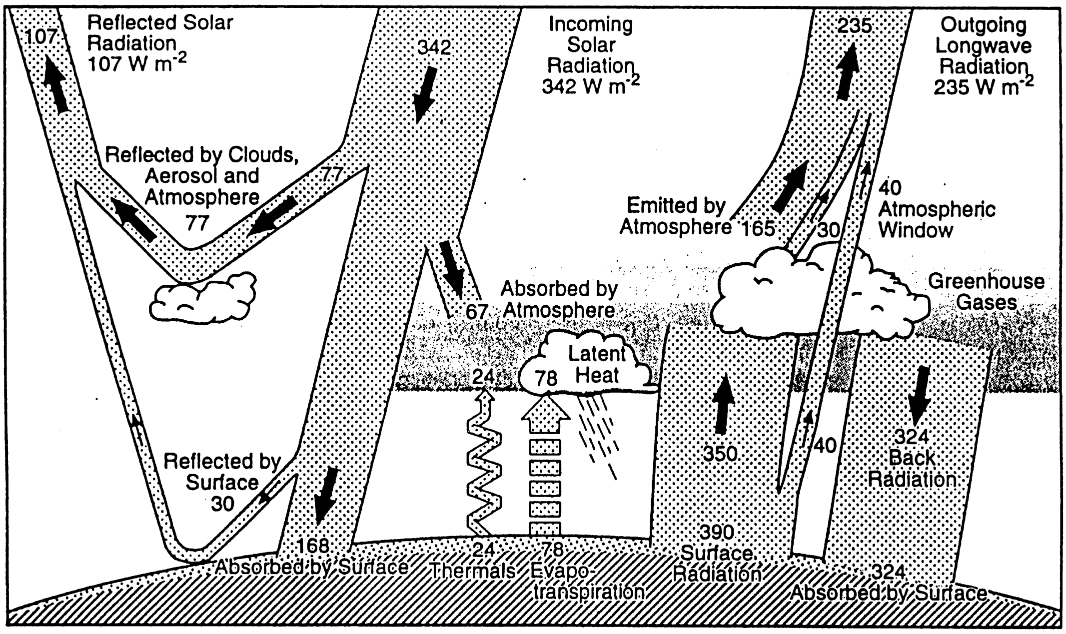

Fig. 3. A detailed sketch of the radiation budget of the Earth–ocean–atmosphere system (from Kiehl and Trenberth, 1997).

It is clear from this that although the effect of an increase in the Sun's total irradiance (which we shall refer to for convenience as the solar constant) at the top of the atmosphere is easily calculated, the effect on temperature and other climatic parameters at the surface is much more difficult to estimate, since changes can take place within the atmosphere–ocean system that can make the response highly nonlinear. For example, one result of an increase in the solar constant might be increased evaporation, tending to cool the surface and warm the troposphere, leading initially to a more vigorous hydrological cycle. Associated with this enhanced hydrological cycle, however, is increased cloudiness, which would tend either to cool or to warm the surface, depending on cloud height and cloud microphysics, as will be discussed later. Estimating the strength of feedbacks of this kind and their regional variations is extremely difficult, accounting for much of the uncertainty surrounding current model predictions of future climate change.

Variations of the solar constant have been used in several general circulation model (GCM) experiments to test model sensitivity and to compare with the effect of greenhouse-gas increases. One of the earliest experiments was that of Wetherald and Manabe (1975), using an early version of the Geophysical Fluid Dynamics Laboratory Global Climate Model (GCM). The solar constant was increased by 2% and decreased by 2 and 4% in separate runs, and the results compared with those from a standard value. Some highly nonlinear effects were found. For example, decreasing the solar constant from -2 to -4% caused the snowline to move equatorward by 5°, and the accumulated snowfall amount to increase by about 30%, an effect that was attributed to the snowline crossing the major midlatitude storm track, and effectively changing rain into snow. Since snow cover causes a large increase in albedo, the change in the surface radiation budget is much greater than the 2% change in incoming solar radiation. The model, however, was a very simplified one, with no seasonal variation and fixed cloudiness, and a more realistic experiment using the National Center for Atmospheric Research Community Climate Model was carried out by Marshall et al. (1994), though again with unrealistically large variations in the solar constant ranging from -5 to $+5\%$. They found a nearly linear dependence of the globally averaged temperature on the solar constant, with the remarkably high sensitivity of nearly 4°C at the surface for a 1% change, compared with the rough estimate above of 0.6°C in the effective equilibrium radiating temperature at the top of the atmosphere. The largest changes in temperature were found at high latitudes in winter, when solar radiation is weakest, illustrating the importance of changes in heat transport. Markedly different results were produced by the Goddard Institute for Space Studies GCM (Rind and Overpeck, 1993) when the solar constant was reduced by the more realistic amount of 0.25%. The global average temperature was reduced by 0.45°C, translating to a sensitivity of only 1.8°C for a 1% change in the solar constant, assuming that a linear extrapolation is valid. In contrast with the Marshall et al. (1994) results, there was very little increase in the temperature at high latitudes, and some regions actually warmed, due to changes in

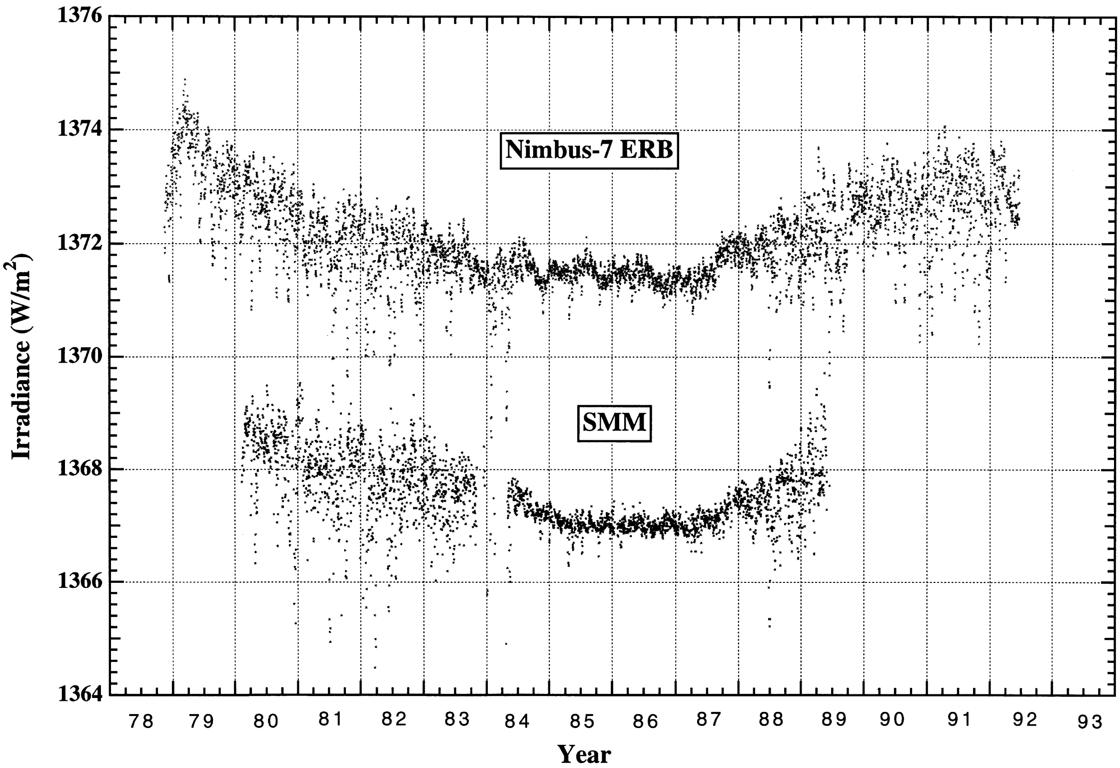

Fig. 4. Daily measurements of the Sun's total irradiance from two spacecraft experiments.

circulation leading to increased advection of warm air. The extent to which the differences between the two model calculations are due to the differences in the size of the solar constant changes or to differences in the models themselves is not clear, and requires further study. A more recent experiment using a model derived from the GISS GCM (Hansen et al., 1997) in which the solar constant was reduced by 2% produced cooling of about 3°C over the $\pm 40°$ latitude range, with greater cooling at high latitudes, reaching over 8°C in polar regions, where sea-ice formation leads to a strong positive feedback through increased albedo.

Turning to observations, the variations in the solar constant that have been measured on the decadal time scale are much smaller than those used in the modeling studies, as mentioned earlier. Figure 4 shows daily values of the solar total irradiance measured by the Active Cavity Radiometer Irradiance Monitor (ACRIM) instrument on the Solar Maximum Mission satellite and the Earth Radiation Budget experiment on the Nimbus-7 satellite. Absolute values differ by about 4 W m^{-2} between the two instruments, but the individual daily departures from the long-term average are highly correlated. Day-to-day variations amounting to a few parts per thousand are associated with the appearance of magnetically active regions on the Sun, and are superimposed on a slower variation that is in phase with the 11-year sunspot cycle, dropping by about 0.1% from sunspot maximum in 1979 to sunspot minimum in 1986. The measurements have continued on a variety of spacecraft, most recently on the Solar and Heliospheric Observatory (SOHO), launched in late 1995, and the decadal-scale component of total irradiance has continued to track the 11-year activity cycle. Careful analysis of sea-surface temperature records by White et al. (1997) has shown the existence of variations in all of the major ocean basins amounting to a few hundredths of a degree that are also in phase with the 11-year solar cycle, and are of the order of magnitude expected as a response to the small 11-year variation in total irradiance. The temperature variations extended down through the upper mixed layer of the ocean, and their existence appears to establish the expected link between irradiance variations and the Earth's climate, although the magnitude is probably too small to explain most of the reported correlations between solar activity and climate.

Following up on this, it is reasonable to ask whether the larger global temperature variations that have been inferred from past climatic variations could also have been caused by larger variations in total irradiance,

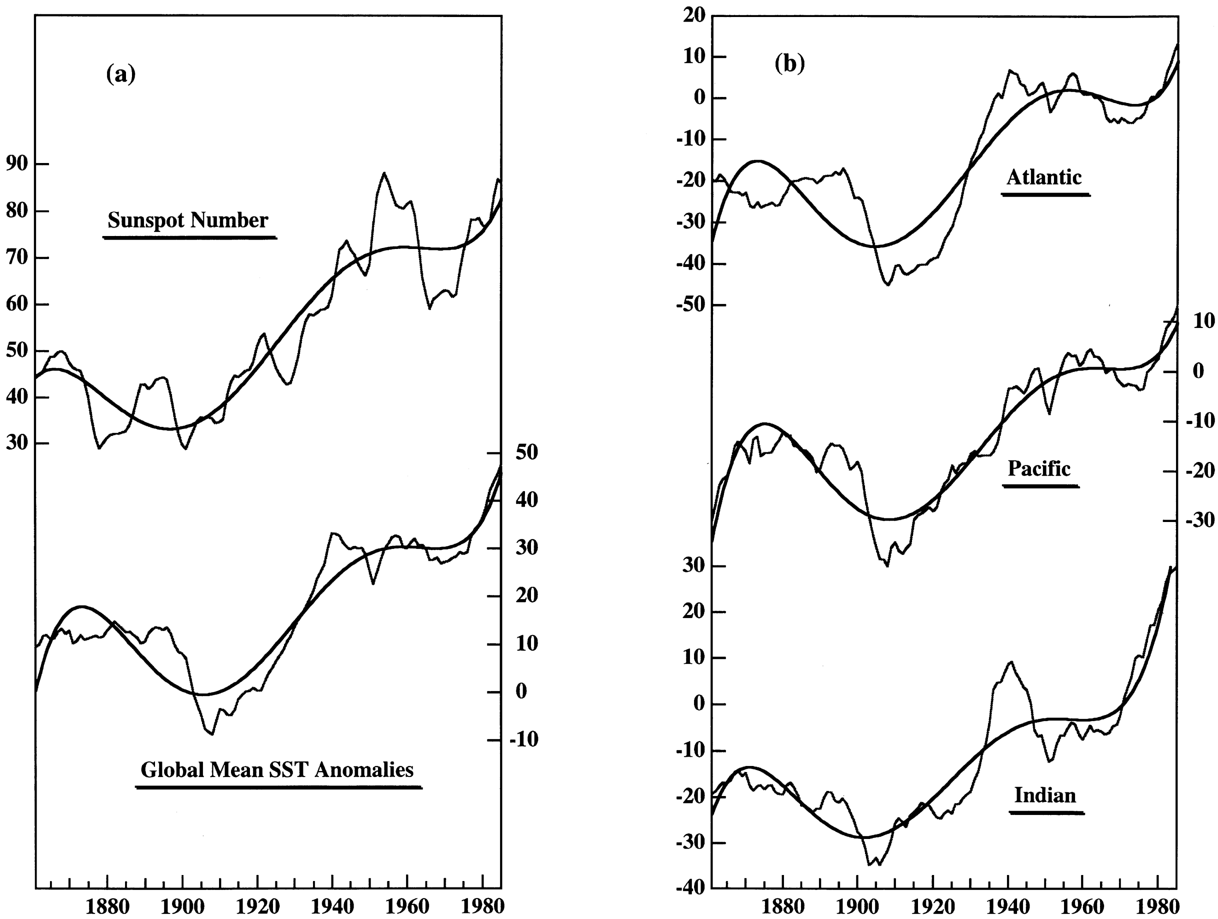

Fig. 5. Eleven-year running mean sunspot numbers and departures of sea-surface temperatures from the long-term mean (units: hundredths of a degree Celsius; data from Bottomley et al., 1990).

occurring on time scales longer than that of the 11-year cycle. Such variations could be related to the overall level of solar activity represented, for example, by the envelope of the 11-year solar cycle, as suggested by Eddy (1975). On the time scale of decades to centuries, global temperatures seem to have varied roughly in step with the average level of solar activity as defined by the sunspot record in recent times, and by the C^{14} record during earlier periods. In particular, Eddy (1975) pointed out the coincidence between the Maunder Minimum of solar activity about 300 years ago and one of the coldest episodes of the Little Ice Age, which gripped the Earth from the early 16th to the late 19th century (Grove, 1988). The average level of solar activity has shown a steady increase since the end of the Little Ice Age, reaching a peak in the late 1950s, but continuing at a high level since then. Global temperatures have also increased, and the temperature record has continued to show long-term variations that match the variations in the overall level of solar activity (Reid, 1991). Figure 5(a) shows 11-year running mean departures of global sea-surface temperatures from the long-term average (Bottomley et al., 1990) together with the 11-year running mean sunspot number, which is an indicator of the overall level of solar magnetic activity. The least-squares polynomial fits shown by the smooth curves have almost identical features, with minima in the early 20th century and weak secondary minima in the 1970s. Figure 5(b) shows that all three major ocean basins contribute to this global temperature signal, indicating that the variation is probably externally forced, and not the result of stochastic climate fluctuations resulting from the strong nonlinearity of the climate system. Neither stochastic fluctuations nor variations resulting from changes in deep-water formation and the global thermohaline circulation, which have also been invoked as a major cause of global climate change, are likely to lead to a similar signal everywhere. Sustained periods of intense volcanic activity could presumably produce a similar variation in global temperature, but it does not appear that this could explain the variation over the past century, when volcanic activity has been generally well documented. If volcanic eruptions were to be the cause, one would also be faced with the remarkable correlation between volcanic activity and the sunspot cycle, which would have to be dismissed as a curious coincidence. Another possible explanation of the similarity in the temperature signal from all the oceans could be changes in instrumentation that might have affected ship-borne measurements worldwide. Such changes have been taken into consideration as far as possible in preparing the sea-surface temperature time series, and the lack of any bias of this kind is also shown by a comparison with a similar time series of nighttime air temperatures made with conventional thermometers (Bottomley et al., 1990). Globally averaged night marine air temperatures are highly correlated with the sea-surface temperatures, with a correlation coefficient of +0.88 and a least-squares straight-line slope of 0.99.

Friss-Christensen and Lassen (1991) have shown a striking correlation between land air temperatures over the northern hemisphere and the length of the sunspot cycle, which is itself correlated with the long-term average of the sunspot number and is thus an alternative proxy for solar magnetic activity. Land and ocean temperatures appear to be displaced in time by several years (Reid, 1991), for reasons that are not entirely clear, as are the time series of averaged sunspot number and solar-cycle length. These time displacements account for the differences in the degree of correlation between land temperatures and the solar-cycle length in one case, and ocean temperatures and the average sunspot number in the other, but both can be taken as strong indicators of the reality of solar forcing of global climate change on the time scale of decades to centuries. Sea-surface temperatures are probably a better proxy for the global temperature than land air temperatures since oceans cover over 70% of the global surface, and sea-surface temperatures are less subject to regional and temporal variation than land-surface temperatures.

Globally averaged temperatures at the peak of the Little Ice Age have been estimated as between 1.0 and 1.5°C lower than modern temperatures (e.g., Crowley and North, 1991). If these low temperatures and the subsequent variations during the more recent warming are to be explained as a result of changing solar irradiance, the changes must be at least several tenths of a percent, and certainly considerably larger than the variation seen by spacecraft to date. Lean et al. (1995) have extrapolated the relationship observed between solar magnetic activity and total irradiance during the past decade to conditions during the Maunder Minimum by using inferences from the behavior of other Sun-like stars (Baliunas and Jastrow, 1990). They concluded that the solar constant could have been about 0.24% lower than its modern value during the Maunder Minimum, and that this would have resulted in a globally averaged temperature about 0.45°C lower than today's value, using the GISS GCM (Rind and Overpeck, 1993). Since the GISS model has a climate sensitivity that is near the upper end of the range of GCM sensitivities (Cess et al., 1990), other models would have produced a smaller global cooling, which is likely to have been too small to explain the low temperatures reached during the Little Ice Age as quoted above. Nevertheless, reconstruction of the total irradiance history based on this estimate suggested that up to about one-third of the global warming that has taken place since the early 20th century could have been caused by solar variability. A different approach was taken by Reid (1997), who assumed that the Maunder Minimum temperature was about 1°C lower than modern temperatures, in rough agreement with the estimates above, and then used a one-dimensional model of the ocean-atmosphere thermal structure to show that this required a solar constant about 0.6% lower than today's value. A reconstruction of solar-constant history on this basis suggested that up to about one-half of the 20th century warming could have been a result of solar forcing. In either case, it appears that solar variability is a factor that should be taken into account in attempting to assess past and future global climate change.

The analysis of Lean et al. (1995) suggested that irradiance variations associated directly with the surface aspects of solar activity (sunspots, faculae, and the chromospheric network) are unlikely to exceed 0.2–0.3% on the time scale of decades to centuries. Unless the climate sensitivity to solar forcing is considerably greater than current models allow, larger variations in irradiance probably occur, but the physical mechanism responsible for these variations remains mysterious. The generation of the Sun's radiation by thermonuclear processes in the core cannot be involved, since the appropriate time constants are of the order of billions of years, but much shorter time constants apply to the transport of radiation from the core to the surface via the convection zone. The convection zone occupies about one-third of the solar radius, and the time required for magnetic flux tubes to rise from the base to the surface is probably several years (e.g., Dicke, 1979). Changes taking place in the convection zone may thus reflect the long-term level of solar activity, providing at least a potential link between solar irradiance variations and the overall level of solar activity. Hoyt and Schatten (1993) reconstructed a time history of irradiance variations since the early 18th century on the basis of a number of parameters that they felt might be proxies for convective transport efficiency, while Nesme-Ribes and Manganey (1992) have used observations related to convection zone dynamics to deduce a Maunder Minimum irradiance about 0.5% lower than the current value. The question of the structure and time-varying behavior of the Sun's convective zone is a major challenge for solar physics, to which the growing science of helioseismology may be able to make an important contribution.

4. Variations in the solar wind

While variations in the solar wind have no direct impact on the surface environment, they give rise to a modulation in the flux of galactic cosmic rays reaching the Earth's atmosphere. The fractional variation associated with the 11-year solar cycle is much larger than the small variations in total irradiance which we have been discussing above, amounting to 15–20% at midlatitudes. Since the effect of the solar wind is to hinder the entry of galactic cosmic rays to the heliosphere, the cosmic-ray flux is anticorrelated with solar activity, and is a strong function of the energy of the incoming particles, and hence of geomagnetic latitude. A convenient proxy for the flux is provided by neutron monitors, which record neutrons generated chiefly by the primary cosmic-ray protons that ionize the lower stratosphere and upper troposphere. Ney (1959) first suggested that the solar-cycle modulation might have an effect on terrestrial climate, possibly through the effect of atmospheric ionization on thunderstorm activity. Even if thunderstorms were affected, however, the relevance to global climate is not obvious, and Dickinson (1975) suggested instead that cosmic-ray ionization would lead to the existence of large ion clusters in the lower stratosphere and upper troposphere that could serve as nuclei for the heterogenous nucleation of sulfate aerosol, and also as condensation nuclei for formation of cloud particles.

Clouds form an important component of the climate system, but an extremely difficult one to deal with in terms of their effect on climate. By reflecting and scattering incoming solar radiation back into space they tend to cool the Earth, but by trapping outgoing longwave radiation from the underlying atmosphere and surface they also tend to heat the Earth. The cloud height is the main parameter determining which of these effects predominates. Low clouds do not trap much longwave radiation and radiate to space at a temperature not much lower than that of the surface, so their chief impact is cooling through increased albedo. High clouds, on the other hand, radiate to space at a much lower temperature, so their longwave trapping efficiency is high and their main impact is to warm the underlying regions, particularly if their optical depth in the visible is small. Recent spacecraft measurements of outgoing and incoming radiation have shown that the net global effect of clouds is to cool the Earth, but warming is dominant in some regions, particularly at high latitudes in winter, when direct solar radiation is relatively weak. The magnitude of the impact of clouds on climate is large. In a one-dimensional climate model simulation that included clouds, Schneider (1972) found that changing the cloud amount by 8% or changing the height of the clouds by 0.5 km were each roughly equivalent to changing the solar constant by 2%, which is a much larger change than those discussed above as being necessary to explain the difference between today's climate and that of the Little Ice Age.

Variations in cloudiness are an attractive means of coupling solar variability to climate, since the formation of a liquid- or ice-phase cloud from water vapor releases latent heat into the atmosphere. A relatively small amount of energy used in creating the conditions necessary for cloud formation can thus release a much greater amount of energy, some of which will be used to heat the atmosphere and some to increase its kinetic energy. The net gain to the Earth-atmosphere system, of course, is zero, since the energy released is equal to that absorbed from the surface by evaporating the water in the first place. The energy is redistributed, however, since the release takes place well above the surface and usually far from the original source. There is a net gain also if the cloud forms in the ice phase, since the further release of latent heat in freezing was not balanced by evaporation from the liquid source.

The lower atmosphere contains enough condensation nuclei in the form of dust particles and other aerosols that liquid-water clouds form when the water-vapor content is only slightly above the saturation value, but this is not the case for ice clouds in the upper troposphere. The crystalline structure of ice imposes fairly rigid conditions on the formation of ice clouds, such as high-altitude cirrus, and cloud particles can exist in a supercooled state down to temperatures of the order of $-40°C$ before freezing. In a series of papers (e.g., Tinsley and Deen, 1991) Tinsley and his colleagues have proposed a mechanism by which changes in the global electric field, driven by solar-wind variations on a wide range of time scales, can stimulate the freezing of supercooled clouds and thereby release latent heat, which can then lead to large-scale changes in tropospheric circulation. While changes in the solar wind do cause changes in the global electric circuit, it has not yet been shown that this 'electrofreezing' mechanism really works, since it depends on two further steps that remain highly speculative. First, electric fields have not been shown to have a well-defined impact on cloud microphysics (although suggestions of an effect have appeared in the literature), and secondly, even if the freezing mechanism takes place, it is not obvious that it will lead to large-scale circulation changes as opposed to small-scale localized effects in the neighborhood of the affected clouds.

The situation regarding the condensation-nuclei scenario of Dickinson (1975) is also not clear. Mohnen (1990) has claimed that ionization effects on the formation of sulfuric acid aerosol in the lower stratosphere are likely to be negligible by comparison with condensation on existing particles of volcanic or meteoritic origin. While upper tropospheric subvisible cirrus clouds have been observed at high latitudes about 20% of the time by the SAGE II satellite instrument (Wang et al., 1996), nothing is known of their relationship to cosmic-ray fluxes. A positive correlation between global cloudiness and the solar cycle, however, has recently been reported by Svensmark and Friis-Christensen (1997), with a latitude dependence corresponding to that of cosmic-ray modulation, so the question remains open, with a clear need for further investigation.

5. Summary and conclusions

Three different aspects of solar variability that could conceivably have a direct effect on the Earth's surface environment have been considered here. The most direct effect of all is produced by variations in the Sun's total irradiance (the solar 'constant'), about 70% of which is absorbed below the tropopause, warming the land and oceans, driving the climate machine, and providing the energy necessary for photosynthesis. The total irradiance is now known to vary at the level of 0.1% on time scales up to that of the 11-year solar-activity cycle, but there is also circumstantial and admittedly speculative evidence for larger variations on longer time scales, perhaps associated with longer cycles of solar activity (e.g., the 80-year Gleissberg cycle). Variations in the range of 0.5–1% on these longer time scales would be sufficient to explain much of the known global climate variability of the last few centuries. While these long-term variations are still hypothetical, the direct effect of the roughly 0.1% variation associated with the 11-year cycle has now been clearly identified in global ocean temperature measurements, and has been seen to extend down to pycnocline depths with an amplitude and phase relative to the irradiance variation of about the expected magnitude (White et al., 1997).

While the small 11-year variation in irradiance can be attributed to the surface features on the Sun associated with active regions (e.g., Foukal and Lean, 1990), the mechanism responsible for larger long-term variations, if they exist, is uncertain. The generation of energy in the Sun's core cannot be involved, since there are solid astrophysical reasons for believing the relevant time constant to be in the billion-year range. Conditions in the outer convective zone of the Sun, accounting for about 30% of its radius, however, can vary on short time scales, as has been pointed out by a number of authors, and in particular the presence of magnetic fields in the convective zone can influence the transport of radiation. Since these same magnetic fields are responsible for the surface activity as they erupt at the photosphere, the potential link between irradiance and the overall level of solar activity is obvious. The generation of magnetic fields by the solar dynamo, the influence of differential rotation on their topology and motions, and their effect on convective transport of radiation are all areas of great uncertainty. Current models of the convective zone suggest that most of the magnetic-field effects on radiation will occur only in the near-surface layers of the convective zone, but so little is known of the properties and behavior of the convective zone that it is hard to say how robust this conclusion is. One parameter that appears to be reasonably secure is the time constant for thermal relaxation of the convective zone as a whole following any transient disturbance, which is about 100,000 years. This is also close to the recurrence period of the glacial episodes on Earth during the Pleistocene epoch, which is usually explained as being a consequence of the variations in the eccentricity of the Earth's orbit. The match in time has never been very satisfactory, however, and the temperature change from glacial to interglacial conditions has seemed too large to be explained by the small changes in eccentricity. Some form of thermal instability in the Sun's convective zone could provide a natural explanation, but one that would be very difficult to establish.

The second category of solar variability involves variations in spectral irradiance. Little is known about spectral irradiance variability in the visible and infrared portions of the solar spectrum, but such variations could have an impact on climate if they exist. The albedos of clouds and the surface are wavelength dependent, and the effective depth in the ocean at which light is absorbed is also a strong function of wavelength. Radiation at the blue end of the spectrum penetrates deeply into the ocean, and can even reach below the base of the mixed layer, where its heating effect will have little impact on the atmosphere. Infrared radiation, on the other hand, is strongly absorbed in the uppermost surface skin of the ocean, and its heating effect can be efficiently transferred to the atmosphere. Until more is known about how the variation in total irradiance is distributed across the spectrum, such effects must remain speculative, and the only impact that can be assessed at present is that of the ultraviolet spectral irradiance. The radiation responsible for stratospheric ozone production ($\lambda \sim 200$ nm) is now known to vary at the level of 3–10% with the 11-year solar cycle, but this translates into a variation of only about 1–2% in total ozone. Both models and observations agree on the size of the total ozone variation, but there is some disagreement on the height distribution, with models emphasizing the middle stratosphere, and observations showing a maximum effect in the lower stratosphere, where the ozone concentration is determined by dynamical effects and not by photochemistry. While this implies that there may be a solar-cycle variation in the dynamics of the lower stratosphere, possibly related to forcing by planetary or gravity waves, more study is needed to establish such a conclusion.

As far as the Earth's surface environment is concerned, the small solar-cycle variation in total ozone implies an equally small variation in surface ultraviolet exposure, probably negligible by comparison with variations caused by anthropogenic and volcanic effects. Modeling experiments suggest, however, that the small changes in UV and stratospheric ozone may have a significant impact on the dynamics of the upper troposphere, and may lead to variations in tropospheric temperatures and storm tracks. Further study is again needed to establish these effects.

The third aspect of solar variability is that of the solar wind, which modulates both the flux of galactic cosmic rays to the Earth's atmosphere and the strength of the global electric field. Electric field variations have been invoked as a potential cause of climatic change through a possible impact on the freezing of supercooled cloud droplets, while cosmic-ray variations may have a significant effect on the concentrations of condensation nuclei in the upper troposphere, and hence on global cloud cover. Neither of these suggested mechanisms has yet been subjected to rigorous quantitative testing or to verification by direct observation, so they remain largely speculative. Recent work, however, has shown what appears to be a highly significant relationship between cloud cover over certain areas of the Earth and the solar cycle for a period of about one and a half cycles. The relationship to either of the above mechanisms is not clear, and more study is needed, in particular to establish the height and other properties of the clouds involved.

In summary, the age-old question of the impact of variations in the Sun on the human environment is far from being settled, but we have come a long way in the last two decades or so. Exciting challenges remain, but the way ahead is difficult and truly interdisciplinary. The clues lie in a wide variety of fields from paleoclimatology through oceanography and meteorology to the plasma physics of the Sun's convective zone. Mastery of all these fields is clearly impossible for a single individual, but a superficial understanding and an awareness of current thinking in each of the fields outside one's own is possible, and may point the way to further progress.

Perhaps the major advance in recent years has been the acceptance of solar variability as at least a potential cause of change in our environment, and not something in the realm of science fiction. There is still a strong Earth-centered focus in the geophysical community, whereby mechanisms of change that are forced within the Earth system are respectable, while those involving some external influence are suspect. The clear evidence that an external influence in the form of the impact of a comet or asteroid gave rise to the Cretaceous–Tertiary extinctions shook the science of paleontology from its firm uniformitarian base, and gave respectability to catastrophism. A more modest, but still significant, change has taken place in the climate community with respect to solar variability as a cause of climatic change, and we are beginning to see a new attitude in which the Earth is regarded as just one of the planets, subject to external influences from the solar system and beyond, and not entirely controlled by its own components. Hopefully, progress will be even more rapid in the decades to come.

References

Baliunas, S., Jastrow, R., 1990. Evidence for long-term brightness changes of solar-type stars. Nature 348, 520–523.

Bottomley, M., Folland, C.K., Hsiung, J., Newell, R.E., Parker, D.E., 1990. Global Ocean Surface Temperature Atlas. U.K. Meteorological Office and Massachusetts Institute of Technology.

Brasseur, G., 1993. The response of the middle atmosphere to long-term and short-term solar variability: A two-dimensional model. J. Geophys. Res. 98, 23,079–23,090.

Brewer, A.W., 1949. Evidence for a world circulation provided by measurements of helium and water vapour distribution in the stratosphere. Quart. J. Roy. Met. Soc. 75, 351–363.

Brown, G.M., John, J.I., 1979. Solar cycle influences in tropospheric circulation. J. Atmos. Terr. Phys. 41, 43–52.

Cess, R.D. et al., 1990. Intercomparison and interpretation of climate feedback processes in 19 atmospheric general circulation models. J. Geophys. Res. 95, 16,601–16,615.

Crowley, T.J., North, G.R., 1991. Paleoclimatology, Oxford University Press, New York.

Dicke, R.H., 1979. Solar luminosity and the sunspot cycle. Nature 280, 24–27.

Dickinson, R.E., 1975. Solar variability and the lower atmosphere. Bull. Amer. Met. Soc. 56, 1240–1248.

Eddy, J.A., 1976. The Maunder Minimum. Science 192, 1189–1202.

Foukal, P., Lean, J., 1990. An empirical model of total solar irradiance variation between 1874–1988. Science 247, 556–558.

Friis-Christensen, E., Lassen, K., 1991. The length of the solar cycle: An indicator of solar activity closely associated with the terrestrial climate. Science 254, 698–700.

Grove, J.M., 1988. The Little Ice Age. Methuen, London.

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

Hansen, J. et al., 1997. Forcings and chaos in interannual to decadal climate change. J. Geophys. Res. 102, 25679–25720.

Hofmann, D.J., Solomon, S., 1989. Ozone destruction through heterogeneous chemistry following the eruption of El Chichón. J. Geophys. Res. 94, 5029–5041.

Hood, L.L., 1997. The solar cycle variation of total ozone: Dynamical forcing in the lower stratosphere. J. Geophys. Res. 102, 1355–1370.

Hoyt, D.V., Schatten, K.H., 1993. A discussion of plausible solar irradiance variations, 1700–1992. J. Geophys. Res. 98, 18,895–18,906.

Kiehl, J.T., Trenberth, K.E., 1997. Earth's annual global mean energy budget. Bull. Amer. Met. Soc. 78, 197–208.

Kodera, K., Chiba, M., Shibata, K., 1991. A general circulation model study of the solar and QBO modulation of the stratospheric circulation during the northern hemisphere winter. Geophys. Res. Lett. 18, 1209–1212.

Labitzke, K., van Loon, H., 1992. Association between the 11-

year solar cycle and the atmosphere. Part V: Summer. J. Climate 5, 240–251.

Labitzke, K., van Loon, H., 1997. Total ozone and the 11-year sunspot cycle. J. Atmos. Solar-Terr. Phys. 59, 9–19.

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

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

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

Marshall, S., Oglesby, R.J., Larson, J.W., Saltzman, B., 1994. A comparison of GCM sensitivity to changes in CO_2 and solar luminosity. Geophys. Res. Lett. 21, 2487–2490.

Mohnen, V.A., 1990. Statospheric ion and aerosol chemistry and possible links with cirrus cloud microphysics—a critical assessment. J. Atmos. Sci. 47, 1933–1948.

Nesme-Ribes, E., Manganey, A., 1992. On a plausible physical mechanism linking the Maunder Minimum to the little ice age. Radiocarbon 34, 263–270.

Ney, E.P., 1959. Cosmic radiation and the weather. Nature 183, 451–452.

Reid, G.C., 1991. Solar total irradiance variations and the global sea surface temperature record. J. Geophys. Res. 96, 2835–2844.

Reid, G.C., 1997. Solar forcing of global climate change since the mid-17th century. Clim. Change 37, 391–405.

Rind, D., Overpeck, J., 1993. Hypothesized causes of decadal-to-century climate variability: Climate model results. Quat. Sci. Rev. 12, 357–374.

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

Rosenlof, K.H., 1995. Seasonal cycle of the residual mean meridional circulation in the stratosphere. J. Geophys. Res. 100, 5173–5192.

Schneider, S.H., 1972. Cloudiness as a global climatic feedback mechanism: The effects on the radiation balance and surface temperature of variations in cloudiness. J. Atmos. Sci. 29, 1413–1422.

Svensmark, H., Friis-Christensen, E., 1997. Variation of cosmic ray flux and global cloud coverage—a missing link in solar-climate relationships. J. Atmos. Solar-Terr. Phys. 59, 1225–1232.

Tinsley, B.A., Deen, G.W., 1991. Apparent tropospheric response to Mev-Gev particle flux variations: A connection via electrofreezing of supercooled water in high-level clouds? J. Geophys. Res. 96, 22,283–22,296.

Wang, P.-H., Minnis, P., McCormick, M.P., Kent, G.S., Skeens, K.M., 1996. A 6-year climatology of cloud occurrence frequency from Stratospheric Aerosol and Gas Experiment II observations (1985–1990). J. Geophys. Res. 101, 29,407–29,429.

Wetherald, R.T., Manabe, S., 1975. The effects of changing the solar constant on the climate of a general circulation model. J. Atmos. Sci. 32, 2044–2059.

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

Willson, R.C., Hudson, H.S., 1991. The sun's luminosity over a complete solar cycle. Nature 351, 42–44.