Solar change and climate: an update in the light of the current exceptional solar minimum

BY MIKE LOCKWOOD1,2,*

1 Department of Meteorology, University of Reading, Reading, Berkshire, UK
2 Space Science and Technology Department, Rutherford Appleton Laboratory, Chilton, Didcot, Oxfordshire, UK

Solar outputs during the current solar minimum are setting record low values for the space age. Evidence is here reviewed that this is part of a decline in solar activity from a grand solar maximum and that the Sun has returned to a state that last prevailed in 1924. Recent research into what this means, and does not mean, for climate change is reviewed.

Keywords: climate change; solar activity; centennial variations

1. Context

The history of science reveals a series of ‘controversies’. These often develop into a state where there is little debate within the relevant academic community (and what there is tends to be on peripheral issues), yet widespread popular debate remains. This usually occurs because the research has challenged the beliefs of a significant fraction of the population at large. The nature of the controversies, however, has changed. Where, for example, the advances made by Galileo and Darwin faced opposition because they challenged the established teachings of organized religion, climate scientists in the developed world have faced opposition from their more secular societies because they challenge beliefs that justify lifestyles and/or political allegiances (Malka et al. 2009; Nisbet 2009). But there is a crucial difference about the climate change debate compared with many of its predecessors: humankind could often afford to wait for previous controversies to abate and the main damage done was an unnecessary delay to the implementation of some of the benefits of the research. There is evidence that public opinion on climate change has changed rapidly in many countries and among many demographic groups (e.g. Staudt 2008; Sampei & Aoyagi-Usui 2009), but there is also evidence that time for effective action is extremely short (Kriegler et al. 2009; Vaughan et al. 2009).

*m.lockwood@reading.ac.uk

One contribution to the 2010 Anniversary Series: a collection of reviews celebrating the Royal Society’s 350th Anniversary.
There is a third new dimension to the public debate about anthropogenic climate change—the Internet. Georg Christoph Lichtenberg (1742–1799) was the first professor of experimental physics in Germany. He was, himself, no stranger to controversy and often satirized the misuse of science. His notebooks from the second half of the eighteenth century are full of comments of remarkable relevance to the role that the Internet has played in the climate science debate. For example, he wrote ‘the most dangerous untruths are truths slightly distorted’ and ‘blind unbelief in one thing springs from blind belief in another’ (Katritzky & Lichtenberg 1984). The Internet has played a useful role in conveying some of the understanding, images and data that lead climate scientists to their conclusions. However, it has also become a haven for un-refereed pseudo-science with dangerously incorrect inference. It has served to give the false impression that there is a serious, widespread academic debate on the basic nature of climate change. The most popular argument runs like this: ‘The Sun drives Earth’s climate system. Therefore changes in the Sun must drive changes in Earth’s climate system’. The first sentence is, of course, absolutely correct; but understanding why the second sentence does not follow from the first requires scientific training and study. The urgency of the need to resolve this distinction between academic and popular understanding places scientists in a dilemma. By trying to find easy-to-understand yet irrefutable ways to demonstrate the fallacy of the above argument (and others that are more intricately constructed but no less misleading), one risks appearing to give scientific credibility and exposure to ideas of no scientific merit.

Lockwood & Fröhlich (2007) demonstrated that since 1987 the long-term changes in solar outputs, which have been postulated as drivers of climate change, have been in the direction opposite to that required to explain, or even contribute to, the observed rise in Earth’s global mean air surface temperature (GMAST). Since then, the solar trends noted by these authors have continued. By the end of solar cycle 23, the annual mean of the open solar magnetic flux (deduced from geomagnetic activity) had fallen to a value last seen in 1924, in the minimum between sunspot cycles 15 and 16 (Lockwood et al. 2009). Other aspects of this decline in solar activity are reviewed by Russell et al. (in press). In this paper, we study the implications.

2. Some basic considerations

The electromagnetic spectrum of the Sun is close to that of a blackbody radiator with effective mean surface temperature \( T_{\text{sun}} = 5770 \text{K} \) (see fig. 1 of Gray et al. submitted). The emitted irradiance from unit area of the surface \( I \) is \( \sigma \varepsilon T_{\text{sun}}^4 \), where the emissivity \( \varepsilon \) is very close to unity and \( \sigma \) is the Stefan–Boltzmann constant. Were the Sun a completely uniform, isotropic radiator, then its total luminosity would be \( 4\pi R_{\odot}^2 I \), where \( R_{\odot} \) is the mean solar radius (6.960 × 10^8 m) and the power falling on unit perpendicular area at the mean Sun–Earth distance (\( R_1 = 1 \text{AU} = 1.496 \times 10^{11} \text{m} \)) would be

\[
I_{TS} \approx \frac{4\pi R_{\odot}^2 I}{(4\pi R_1)^2} \approx \sigma T_{\text{sun}}^4 \left( \frac{R_{\odot}}{R_1} \right)^2.
\]  

(2.1)
This is the total solar irradiance (TSI). This relation is approximate because the emission is not, in general, uniform (for example, there are limb-darkening effects) nor isotropic (for example, facular brightening is at low elevation angles and faculae are not evenly distributed across the solar disc). For the surface temperature $T_{\text{sun}} = 5770$ K, equation (2.1) gives $I_{\text{TS}} = 1360 \, \text{W m}^{-2}$. Composites of measurements, related to a standard source, have given mean values $\langle I_{\text{TS}} \rangle$ at Earth near $1365.5 \, \text{W m}^{-2}$ (e.g. Fröhlich 2006). However, data from the latest instrument, the total irradiance monitor (TIM) on the SORCE satellite launched in 2003, are yielding very similar variations, but values are persistently lower (see §5 and figure 2) with $\langle I_{\text{TS}} \rangle = 1361 \, \text{W m}^{-2}$ (Kopp et al. 2005). The origins of this discrepancy are not yet clear, and resolution will ultimately require future instrumentation. In particular, the Glory mission is due for launch in October 2010 and will carry Glory/TIM, which will have had end-to-end testing and calibration using the new TSI radiometer facility.

The global mean solar ‘shortwave’ (SW) power input to Earth’s climate system per unit surface area, $P_{\text{in}}$, is given by:

$$P_{\text{in}} = I_{\text{TS}} \frac{(1 - A)}{4} = \varepsilon_{\text{E}} \sigma T_{\text{E}}^4 + \gamma = \sigma T_{\text{S}}^4 - G + \gamma,$$

where $A$ is the Earth’s global albedo, our best estimate of which is very close to 0.3 (Loeb et al. 2009). The factor 4 is the ratio of the area of the disc presented to the Sun by the Earth to its total surface area. The right-hand term of equation (2.2) is the mean of the infra-red (‘longwave’) power that Earth radiates back into space plus the mean power absorbed by the planet: $T_{\text{S}}$ is the Earth’s GMAST, $T_{\text{E}}$ is the ‘effective radiation temperature’ of the Earth, as seen from space, $\varepsilon_{\text{E}}$ is the emissivity of the Earth ($\varepsilon_{\text{E}} = 1$ only if the Earth can be approximated to a blackbody radiator) and $G$ is the total radiative forcing due to greenhouse gases (GHGs) and other causes not included in $A$ ($G = \sigma T_{\text{S}}^4 - \varepsilon_{\text{E}} \sigma T_{\text{E}}^4$). The term $\gamma$ makes allowance for the fact that the Earth is not quite in radiative equilibrium. For $A = 0.3$ and $I_{\text{TS}} = 1361 \, \text{W m}^{-2}$, equation (2.2) gives $P_{\text{in}}$ of $238 \, \text{W m}^{-2}$, and out of this Hansen et al. (2005) estimate from observed ocean temperature rises that $\gamma = 0.85 \pm 0.15 \, \text{W m}^{-2}$ is entering and warming the oceans. Although this is a very small fraction of $P_{\text{in}}$, it is important because of its contribution to the ‘temperature commitment’ and means that even if GHG emissions were restricted to stabilize $G$ at current values, $T_{\text{S}}$ would still rise by a further $0.6 \, ^\circ\text{C}$. Subsequent investigations, using both models and observations, have confirmed that the value of $\gamma$ given by Hansen et al. (2005) is broadly correct (Murphy et al. 2009; Trenberth et al. 2009).

There are some simple calculations that can be made using equation (2.2) for the $A$ and $I_{\text{TS}}$ values given above. If the greenhouse forcing $G$ were zero (and $\gamma = 0$), equation (2.2) gives $T_{\text{S}} = 254 \, ^\circ\text{C} = -19 \, ^\circ\text{C}$. Raising this to Earth’s average pre-industrial temperature (taken here to be an average of estimates for 1650–1700, Juckes et al. 2007) of $T_{\text{S}} = 287 \, ^\circ\text{C}$ (= $14 \, ^\circ\text{C}$) requires $G = 146.5 \, \text{W m}^{-2}$. Raising this further by a conservative estimate of the temperature anomaly rise since 1700 of $\Delta T_{\text{S}} = 0.8 \, ^\circ\text{C}$ (with $\Delta \gamma = 0.85 \, \text{W m}^{-2}$ and constant $A$ and $I_{\text{TS}}$) requires an increase in $G$ of $\Delta G = 5.15 \, \text{W m}^{-2}$.

Section 5 discusses the solar cycle variation in TSI: its peak-to-peak amplitude is $1.37 \, \text{W m}^{-2}$, which, in the absence of any amplification mechanism, corresponds to a radiative forcing change of $0.24 \, \text{W m}^{-2}$ (for $A = 0.3$). This is less than
5 per cent of the required total change in the radiative forcing since pre-industrial times derived above \((\Delta G = 5.15 \, \text{W m}^{-2})\). To explain the observed GMAST rise since 1700 would require a change in TSI of 29.1 W m\(^{-2}\). Detection–attrition studies using climate models have found a feedback amplification of the input centennial TSI variation by a factor of \(\beta_S \approx 3\) (e.g. Stott et al. 2003; Ingram 2006), which would reduce the required TSI change to near 10 W m\(^{-2}\). However, one should note that some of this amplification may be unreal and due to a metrology factor (discussed in §9), which has been reported as exaggerating the downward drift in GMAST in the middle of the twentieth century. The inferred \(\beta_S\) factor is particularly sensitive to this because the mid-century downward drift is partly explained by the solar forcing variation. Note that if the mean effective temperature of the solar surface has remained constant (and with \(\beta_S = 1\)), the simplified equation (2.1) shows that the solar radius \(R_\odot\) would have to have been 1.1 per cent smaller a century ago than it is today to supply the required radiative forcing change.

From flask measurements and ice-core data, we can compute the radiative forcing caused by the observed rises in well-mixed trace (i.e. excluding water vapour) GHG abundances over the same interval (Forster et al. 2007). The abundance of carbon dioxide has risen from 280 p.p.m.v. in pre-industrial times to 362.5 p.p.m.v. in 2000, which corresponds to \(\Delta G\) of \(+1.56 \, \text{W m}^{-2}\). The corresponding radiative forcing contributions for other trace GHGs are: \(+0.47 \, \text{W m}^{-2}\) for \(\text{CH}_4\), \(+0.28 \, \text{W m}^{-2}\) for chlorofluorocarbons, hydrochlorofluorocarbons and halons and \(+0.14 \, \text{W m}^{-2}\) for \(\text{N}_2\text{O}\). This gives a total of 2.45 W m\(^{-2}\). This is close to half of the required forcing, most of the remainder being explained in climate models by feedback loops.

We know very little about variations in Earth’s albedo, \(A\), and almost nothing on centennial time scales (Charlson et al. 2005): in order to cause the observed GMAST rise of \(\Delta T_S = 0.8^\circ\text{C}\), with no change in TSI or \(G\), it would require Earth’s albedo \(A\) to have fallen by 5 per cent from 0.315 to 0.300. This is an average rate of change (over the 300-year interval) of \(dA/dt = -5 \times 10^{-4} \text{yr}^{-1}\). Recent re-analysis by Pallé et al. (2009) has improved the degree of agreement between ISCCP/FD and CERES satellite data and the Earthshine (lunar Ashen light) method. For the ISCCP/FD and Earthshine data, they found that the albedo has increased over the interval 1999–2005 at a rate of about \(dA/dt = +3.5 \times 10^{-4} \text{yr}^{-1}\). The CERES data re-analysis shows no detectable change, within limits of approximately \(\pm 5 \times 10^{-5} \text{yr}^{-1}\). Thus recent changes are smaller than the average required over the past 300 years, and the reported changes have been in the opposite direction to that required to explain a rise in GMAST.

During the current solar minimum, the Sun has set many record lows (Russell et al. in press). Values of many solar and heliospheric parameters have moved outside the range that had been directly measured before (figure 1). However, for space measurements of the Sun, the term ‘since records began’ refers only to since the mid-1960s (the ‘space age’). Figure 1 shows that record lows have been set in the PMOD TSI composite (Fröhlich 2009) and in the open solar magnetic flux (Lockwood et al. 2009a,b). Galactic cosmic rays (GCRs) reaching Earth’s atmosphere at high latitudes (where the low-energy part of the GCR spectrum is not excluded by the geomagnetic field) are showing record or near-record highs: figure 1 shows the count rate from the neutron monitor at Oulu, Finland, which is recording greater GCR fluxes than at any time in its history: lower-latitude
stations are showing GCR fluxes that are equalling the previous record high (Moraal et al. 2009). In addition, the long sequence of BiSON helioseismology data is revealing larger changes in the solar interior than seen in previous minima (Broomhall et al. 2009). In this review, we re-evaluate the implications of the data from this low solar minimum.

3. Magnetic field effects on heat flux in the solar convection zone

In recent years, the technique of helioseismology has revolutionized our view of motion of plasma inside the Sun and the consequent generation of magnetic field (see reviews by Gizon & Birch 2005; Hill 2008; Howe 2009). It has long been known that the energy developed in the core of the Sun is transferred across the radiative zone (RZ) by radiative transfer, but then is transferred to the solar surface (the photosphere) because of large-scale motions in the convective zone (CZ) above it. The RZ–CZ boundary sits at a depth below the solar surface of \( z \approx 1.2 \times 10^5 \text{ km} = 0.17R_\odot \), set by the onset of the convective instability. This instability is analogous to how unstable air in Earth’s atmosphere rises in thunderstorms (in fact, plasma ionization changes in the solar plasma play the role that water vapour
condensation does in the atmospheric case). As the CZ fluid rises, it expands and cools; the temperature falling from $2 \times 10^6$ K at the RZ–CZ boundary to $T_{\text{sun}}$ at the surface.

Magnetic field in the CZ is in the form of flux tubes, which can interfere with the convective rise of energy in two ways. They can partly block the flow of heat (the ‘alpha effect’), and if the field strength $B$ rises/falls, the energy stored/released as magnetic energy density ($B^2/2\mu_0$) acts as a heat flux sink/source, respectively (the ‘beta effect’). Any effects on TSI are called ‘shadow’ effects. In addition, the magnetic pressure changes could, in theory, combine with effects of turbulence to change the radius of the photosphere, $R_\odot$. But, magnetic fields cannot reside long in the CZ as their buoyancy makes them inherently unstable and they rapidly rise to the surface (see review by Fan 2004). The rise time from the base of the CZ to the surface $\tau_R$ varies with the flux tube characteristics, but is typically of the order of 100 days. Helioseismology has also revealed the tachocline, a strong gradient in the flow of the solar interior at which the solid rotation of the core and (most of) the RZ changes into the differential rotation of the CZ. Because of these relatively short rise times, the CZ is unable to give long-term storage of the magnetic field. Hence, modern models of the solar dynamo rely on the velocity shear at the tachocline to achieve toroidal field amplification, and the field is stably stored in an ‘overshoot layer’ just below the RZ–CZ transition (Nandy 2003; Charbonneau 2005). Only when the stored field gets too large does it become unstable and rise rapidly through the CZ to give surface magnetic features (such as sunspots and faculae). Another discovery of helioseismology, meridional circulation of the CZ, lies at the heart of recent modelling of the solar cycle oscillations of the mean field dynamo (e.g. Nandy & Choudhouri 2002). Early indications are that the recent long and low solar minimum has been associated with unusually slow equatorward migration of the deep band on enhanced zonal flow that in previous cycles has been seen prior to the onset of flux eruption to the solar surface (Hill 2008; Phillips 2009).

In addition to $\tau_R$, we need to consider two other important time scales in the CZ. The first is the thermal time scale $\tau_T$ (sometimes called the Kelvin–Helmholtz time scale). This is the time for the temperature of the CZ as a whole to heat up or cool down. If we consider $\tau_T(z)$ to be the time scale for the part of the CZ above a depth $z$, we find that $\tau_T(z) \approx 1$ h at $z = 1000$ km (just below the surface) rising to $10^5$ years for the base of the CZ ($z = 1.2 \times 10^5$ km), because of the large mass of the overlying material. On the diffusive time scale, $\tau_D(z)$, differences in entropy, and so heat flow, between different parts of the CZ can be changed. For $z = 1000$ km, $\tau_D(z) \approx 15$ min, whereas at the base of the CZ, $\tau_D(z) \approx 1$ yr. Thus at the surface, $\tau_T$ and $\tau_D$ are of a similar size, but for increasingly deep layers, the thermal time scale increases rapidly so that $\tau_T/\tau_D \gg 1$. This means that changes below the surface on time scales below $10^5$ years do not occur in thermal equilibrium and so any upward heat flux that is blocked is simply stored in the CZ (see review by Foukal et al. 2006). The time constants described earlier were obtained from calculations in which the CZ was modelled using turbulent diffusion approximations. Magnetohydrodynamical simulations, which accurately reproduce observations of solar convection near the photosphere, confirm the conclusions drawn here by considering the huge thermal inertia of the CZ and explain why magnetic effects from deeper in the CZ have not been observed (Rosenthal et al. 1999).

Proc. R. Soc. A
Hence, our modern understanding of the CZ shows that the large-scale magnetic field of the Sun is stored at the base of the CZ, and any effects this stored field has on heat flow will change the surface temperature (and thus the solar irradiance) only on time scales of the order of $10^5$ years. Magnetic field becomes unstable in this store when field strengths become too large (which happens most at the peak of the 11-year sunspot cycle) and rises rapidly through the CZ and becomes manifest as surface magnetic field on time scales of $\tau \approx 100$ days. A thermal time scale of $\tau_T(z) = 100$ yr corresponds to a depth $z$ of approximately 3400 km. Thus, to give variations in the surface temperature and irradiance on century time scales, subsurface magnetic heat blocks must be at $z \leq 3400$ km, i.e. in the top 1.7 per cent of the CZ. Note that helioseismology reveals that sunspot fields extend down to much lower than this $z$ (e.g. Couvidat et al. 2004; Kosovichev 2004; Gizon & Birch 2005). There is, therefore, no reason to expect the fields at $z = 3400$ km to be significantly different from those at the surface, and so on 100-year time scales we should not expect to see magnetic shadow effects beyond those accounted for by surface magnetic features (§5).

In addition to the strong, mean-field, solar dynamo, a second ‘weak’ dynamo is thought to generate smaller flux tubes (Cattaneo & Hughes 2001). There is a spectrum of spatial scales of surface magnetic features, ranging from sunspots (which vary on the 11-year solar cycle) down to features smaller than instrument resolution (which hardly vary in occurrence). Parnell et al. (2009) used SoHO/MDI and Hinode/SOT data to show that all features follow a power-law distribution of fluxes, produced by processes such as fragmentation, coalescence and cancellation of the fields generated by the strong dynamo.

4. Solar radius changes

In addition to their blocking effect on convective heat flow, the magnetic field in, and just below, the CZ could modulate the solar radius $R_\odot$ and, by equation (2.1), TSI. A great many studies over many years have searched for such variations, both over recent solar cycles and, using eclipse data, on centennial time scales. These studies are inconclusive and often contradictory: it is not clear that any changes detected are bigger than the experimental uncertainties. For example, there are a number of reports that show that $R_\odot$ increases with solar activity (e.g. Chapman et al. 2008), whereas other measurements show the reverse or no consistent change (Lefebvre & Kosovichev 2005; Sofia et al. 2005; Egidi et al. 2006). Bruls & Solanki (2004) demonstrated that surface luminosity features can perturb many techniques, and the results from helioseismology show small and inconsistent changes (Lefebvre et al. 2007).

The best observations place an upper limit of approximately 0.01 per cent on variations in $R_\odot$ over decades (Kuhn et al. 2004; Lefebvre et al. 2006; Djafer et al. 2008), and reports of larger changes remain controversial (e.g. Lefebvre & Kosovichev 2005; Noël 2005). The theory predicts that $R_\odot$ will increase with decreased solar magnetic field (Stothers 2006), and this accords with some of the best observations. This gives the wrong sense of change to contribute to the centennial GMAST rise, and the observed solar cycle changes are at least two orders of magnitude too small to explain the observed magnitude of the GMAST change (as discussed in §2). As discussed in the next section, surface
magnetic features can account for all the observed solar cycle variation in TSI and so there is evidence that solar radius changes do not contribute to a detectable extent.

5. Solar surface effects

Magnetic fields near the solar surface act as heat blocks. If the flux tube diameter is large enough, the blocked heat causes a drop in surface temperature and emission (the blocked heat being stored in the CZ). These are sunspots. For smaller diameter flux tubes, radiation from the walls maintains the temperature and the magnetic pressure means that the density must fall. This means that emission from lower, hotter layers can escape. Viewed from vertically above the surface, these small flux tubes are dark compared with the field-free surface, but viewed from low elevations (i.e. for such features near the edge of the solar disc), they allow additional energy release. These are faculae (see reviews by Lockwood 2004; Foukal et al. 2006). Because the number of smaller flux tubes vastly outweighs the large ones (Parnell et al. 2009), the net effect is to make the TSI larger by 0.1 per cent at sunspot maximum than at sunspot minimum. The SATIRE (Spectral and Total Irradiance Reconstruction) modelling of the net effect of all surface magnetic features, as seen in magnetograph data, explains at least 80 per cent of the observed TSI variation with only one free fit parameter, the pixel filling factor of faculae (see recent reviews by Krivova & Solanki 2008; Domingo et al. 2009).

However, there has been debate about the TSI observations. The problem is that it is necessary to make a composite of the data from several different instruments to look at the decadal-scale variations. In addition, the instruments degrade and some have been self-calibrating, whereas earlier instruments require retrospective re-calibration. The three available TSI data composites are termed: PMOD (Fröhlich 2006), ACRIM (Willson & Mordvinov 2003) and IRMB (Dewitte et al. 2004). Figure 2 shows how the various observations of TSI over the recent decline from solar maximum to solar minimum compare. Figure 2a shows that there is considerable difference between data used in the three composite datasets, but those used by the PMOD and IRMB composites are indeed very similar to the TIM/SORCE data (black points and grey triangles, respectively) other than, as discussed in §2, TIM gives consistently lower values than PMOD and IRMB by 4.4 and 5.4 W m\(^{-2}\), respectively (cf. solid lines). There is considerable scatter in the ACRIM data compared with the others. The smaller values in figure 2a are generated by large sunspot groups passing over the solar disc, and so to look at any difference in trends over a longer time scale, figure 2b displays the 27-day running means along with linear regression fits as dot-dashed lines. The PMOD composite shows a decline that is approximately 11 per cent larger than that in SORCE/TIM data and 9 per cent larger than that in IRMB, but 16 per cent smaller than that in the ACRIM data (which retains the large scatter for these solar rotation means).

All composites have to use data from the early HF instrument on Nimbus 7 during what is termed the ‘ACRIM-gap’. As reviewed by Lockwood & Fröhlich (2008), the ACRIM composite assumes the HF data are adequate in largely uncorrected form during the ACRIM gap, whereas the PMOD composite allows
Figure 2. Scatter plots of various TSI estimates compared with the PMOD composite values for the fall from the last solar maximum to the current solar minimum (all coincident data from February 2003 to September 2009 are used). At this time, the composites all come from a single data source and so intercalibration between successive instruments is not an issue: PMOD comes from a combination of the two SoHO VIRGO radiometers (PMO6V and DIARAD), IRMB comes from SoHO VIRGO/DIARAD only and ACRIM comes from ACRIM-III on ACRIMsat. (a) is for daily means and (b) for 27-day averages. In both cases, the comparison with SORCE/TIM data is shown by filled dots, with the ACRIM composite by open squares and the IRMB composite with grey triangles. Both plots also show equal ordinate and abscissa with a dashed line and solid lines offset by $+1$ and $-4.4\text{Wm}^{-2}$. (b) The dot-dashed lines show linear regression fits to the data (which assume uncertainties in both parameters and minimize the r.m.s. perpendicular deviation of points to the best fit line), which have slope values of 0.889, 0.913 and 1.160 for SORCE/TIM (filled circle), IRMB (grey shaded) and ACRIM (open squares), respectively.

for a known pointing direction glitch. As a result, the ACRIM composite shows a rise in TSI by $0.6\text{Wm}^{-2}$ between the solar minimum between sunspot cycles 21 and 22 (minimum 21/22) and the next solar minimum (22/23), whereas the PMOD composite does not. Wenzler et al. (2009) used the SATIRE reconstruction based on magnetograph data to show that only the PMOD composite agrees with the effect of magnetic surface features only. Thus the additional drift of the ACRIM composite requires subsurface effects or solar radius changes. From equation (2.1), the change in $R_\odot$ needed for the ACRIM composite to be correct would be over 0.02 per cent, which exceeds the upper limit of 0.01 per cent (for the full solar cycle amplitude) placed by the best solar radius measurements. A confusion in this debate arises from semantics and the use of the word ‘reconstruction’. Wenzler et al. (2009) used the SATIRE reconstruction based

Proc. R. Soc. A
on ground-based magnetograph data that extend back to 1974: they therefore used the best and most relevant data on facular features on the Sun that we have. Scafetta & Willson (2009) claim to have resolved the ACRIM-gap debate (in favour of the ACRIM composite) using the SATIRE reconstruction of Krivova et al. (2007). As pointed out by Krivova et al. (2009), this is not an appropriate use of the Krivova et al. (2007) reconstruction, which was devised to allow irradiance to be modelled back to 1700, long before the advent of magnetographs. To do this, Krivova et al. (2007) had to effectively relate facular occurrence to sunspots, constrained by the open solar flux. One should not apply this approximate reconstruction based on a proxy (and designed to give the best possible TSI reconstruction in the era before magnetograph data became available) to the ACRIM gap (for which magnetograph data are available). As discussed in §9, other work on Earth’s response has been based on the ACRIM composite, but the best analysis, using the best information we have, is that by Wenzler et al. (2009) and that does not agree with the ACRIM composite.

Figure 3 studies the long-term TSI trends inherent in the ACRIM, IRMB and PMOD data composites. Solar-cycle-averaged means are derived using the procedure of Lockwood & Fröhlich (2007), which evaluates the variation of the length of the solar cycle $L$, centred on each month considered. Means are then taken over the interval $L$ for each month. Figure 3b shows the results for the sunspot number. As discussed by Lockwood & Fröhlich (2007) and Lockwood et al. (2009a), figure 3b shows that 1985 marks the peak of the long-term variation in solar activity (as quantified by sunspot number) and the current low solar minimum is part of the subsequent decline (as it contributed to the means shown in figure 3 towards the end of the interval). The thin line in figure 3a shows the corresponding IRMB composite means (from which the early Nimbus HF
data have been omitted as the degradation of this non-self-calibrating instrument was rapid at this time). The trend is upward and so anticorrelated with the sunspot number. The variation with the corresponding means of \( R \) is linear and extrapolated to \( R = 0 \) giving a Maunder minimum (MM) value of the TSI that is 1.2 W m\(^{-2}\) higher than today’s value. This runs counter to most expectations of solar behaviour during the MM and European ‘little ice age’. The trend in the ACRIM composite is the grey line in figure 3a, and it can be seen that this bears almost no relation to the sunspot number in figure 3b. In contrast, the PMOD composite shows a variation similar to the sunspot number, although the variation is not linear. The PMOD composite, therefore, suggests a reduced TSI during the MM. Of the three composites, PMOD contains the largest number of detailed corrections for instrumental performance and stability and is the only one to agree with the SATIRE analysis of magnetograph data, which accounts for surface effects (Wenzler et al. 2009).

Lockwood & Stamper (1999) and Lockwood (2002) noted a good correlation between TSI and the open magnetic solar flux \( F_S \) over recent solar cycles and that this provided an explanation of why cosmogenic isotopes were so well anticorrelated with TSI (the fluxes of GCRs, which generate the cosmogenic isotopes being highly anticorrelated with \( F_S \)—see further discussion by Lockwood 2006). However, these authors were aware that correlation over the solar cycle did not mean that the centennial variations would necessarily also correlate and that long-term reconstruction of TSI could not yet rely on this correlation alone (e.g. Foukal et al. 2006). The modelling of Wang et al. (2005) and the reconstruction by Krivova et al. (2007) provided the missing scientific link by looking at the connection of \( F_S \) to the total photospheric flux and thereby constraining the relationships between sunspots and smaller flux tubes. In this way, the sunspot data and the open flux variation inferred from geomagnetic activity data (see review by Lockwood et al. 2009a) can be combined to reconstruct the centennial variation of TSI. These reconstructions give smaller centennial variations than those that had been derived previously from the now-discounted stellar analogues (see review by Lockwood 2004). For example, Krivova et al. (2007) find that the irradiance increases since the MM was between 0.9 and 1.5 W m\(^{-2}\), with a best value of 1.3 W m\(^{-2}\). This gives a radiative forcing change of 0.22 W m\(^{-2}\) (for albedo, \( A \), of 0.3), which is only 4 per cent of that needed to explain the observed GMAST rise (without amplification, i.e. \( \beta_S = 1 \)).

Figure 2 shows that the recent solar minimum has seen a marked drop in TSI in the PMOD composite based on data from the instruments on the SoHO spacecraft and, although the drop in values from TIM on SORCE and in the IRMB composite is not quite as large, it is similar. It has also produced record low values in the modulus of the radial component of the interplanetary magnetic field (IMF), \(|B_r|\) (figure 1), and hence in the (signed) open solar flux, \( F_S \). Lockwood & Owens (2009) have used widely separated spacecraft in the heliosphere to show that, once an appropriate correction for the effects of solar wind speed variability has been made, the relationship \( F_S = 2\pi R^2 |B_r| \) is accurate to within 2 per cent. Fröhlich (2009) has shown that although we have only three independent solar minimum points, with a semi-empirical fourth (for \( F_S \) in the range \( 2.26 \times 10^{14} \) to \( 3.40 \times 10^{14} \) Wb), they do lie very close to a straight line given by \( I_{TS} \) (in W m\(^{-2}\)) = \( (1364.64 \pm 0.40) + (5.34 \pm 2.39) \times (F_S \text{ in } 10^{13} \text{ Wb}) \). There is considerable evidence that the Sun’s magnetic cycle continued during the MM, and there are grand
solar minima in which cosmogenic isotopes rose to higher values than at that time (Steinhilber et al. 2008). However, we can set a minimum to the TSI in the MM by putting $F_S$ to zero. This gives a best estimate of the lower limit in the MM of $I_{TS} = 1364.64 \pm 0.40 \text{ W m}^{-2}$. The optimum value is $0.64 \text{ W m}^{-2}$ less than that observed in the current minimum and $1.24 \text{ W m}^{-2}$ less than the average over the last two solar cycles. This is very similar to the value recently derived by Steinhilber et al. (2009) from the analysis of cosmogenic isotope data and is within the uncertainty range found by Krivova et al. (2007). This supports the concept used by Krivova et al. (2007), namely that all changes in irradiance were associated with surface magnetic fields only.

6. Spectral irradiance changes and SE effects

Both experimental and theoretical estimates of the fraction of the variation in TSI that is due to the UV vary (Pagaran et al. 2009; Withbroe 2009). A much-used model of the UV variability, based on the observations by the Upper Atmosphere Research Satellite, gives the 200–400 nm range as contributing approximately $0.1 \text{ W m}^{-2}$ to the solar cycle TSI variation (Lean et al. 1997). Recent results of the decline to the current solar minimum from the Spectral Irradiance Monitor (SIM) on SORCE give a radically different picture, with $1 \text{ W m}^{-2}$ coming from this range and upward and downward trends in other wavebands cancelling to a large extent and giving a net contribution of only $0.3 \text{ W m}^{-2}$ (Harder et al. 2009). It is not yet clear to what extent this unexpectedly large change reflects different instrumentation or how much it may be due to the long-term trends that have combined with the solar cycle to generate the current unusual solar minimum. The SATIRE modelling of the variability of the spectrum, extended into the UV by Krivova et al. (2006), shows some similarities and some difference to the SIM results on both solar cycle and solar rotation time scales (Unruh et al. 2008), but does not give as large a change in the UV as reported from SIM data.

Almost all of the UV radiation is absorbed in the stratosphere. In particular, wavelengths below 300 nm alter stratospheric chemistry and control production and destruction of ozone. Effects of the associated stratospheric heating in the underlying troposphere, such as slight modulation of the separation of the jet streams, have been noted and modelled (Haigh 2007; Simpson et al. 2009), and recent work is beginning to help explain how non-linear radiative and dynamical coupling may work via perturbation of internal modes of climate variability such as the El-Niño–Southern Oscillation, ENSO, and the North Atlantic Oscillation (Lean & Rind 2008; Rind et al. 2008; Simpson et al. 2009). Foukal & Bernasconi (2008) and Foukal et al. (2009) have studied historic images of the chromospheric features responsible for much of the UV variation and argue that there is no long-term change. However, the images are far from a homogeneous dataset, and this conclusion does not agree with the studies by Tlatov et al. (2009) and Ermolli et al. (2009). Hence, long-term change of the solar UV remains likely, if poorly understood.

Energetic particles, for example, those generated at the shock fronts associated with transient events in interplanetary space, are known to cause ozone depletion at high latitudes, mainly through down-draughted NO$_x$ generated in the mesosphere and thermosphere. These transient events also drive
geomagnetic activity, giving a connection between polar ozone depletion events and geomagnetic activity. In an exciting new development, Seppälä et al. (2009) have reported statistically significant differences in wintertime polar surface air temperatures between years with high and low means of the geomagnetic Ap index. The changes occur in both hemispheres and are of the order of ±5 K. These results agree with previous model predictions of the effects of NO$_x$ generated by energetic particles, which suggest that the surface air temperature changes follow the ozone depletion via circulation changes including a stronger polar vortex (Rozanov et al. 2005).

7. Direct GCR effects

The suggestion that cosmic rays can modulate the formation of some clouds (see review by Kirkby 2007) would influence equation (2.2) through effects on the albedo, $A$, and via the greenhouse term, $G$. The balance of these two effects depends on the altitude of the clouds: the greenhouse trapping effect being the greater for high (colder) clouds and the albedo effect being larger for low clouds. The rise of open solar flux since the MM (§8) has been inferred to give the long-term decline in GCRs, as shown by the cosmogenic isotopes, for example, in the recently published new ice core data (Berggren et al. 2009). It has been postulated that these GCRs increase low-altitude cloud for which the dominant effect on the energy balance is due to the change in the albedo. The decline of these GCRs since pre-industrial times would, if GCRs really can promote the production of low-altitude cloud, be a contribution to temperature rise. The rise in albedo derived by Pallé et al. (2009) from ISCCP satellite data and the Earthshine measurements over 1999–2006 is therefore consistent with this idea, given that mean cosmic ray flux recovered over this interval (following the long-term minimum shortly after the peak of the current grand solar maximum in 1985) (Lockwood & Fröhlich 2007, 2008). However, Pallé et al. (2009) found no change in albedo from the CERES satellite data.

Svensmark et al. (2009) have recently published a superposed-epoch study reporting large (up to 7%) global cloud cover decreases, as detected by a number of satellites, following Forbush decreases in GCR fluxes. The difficulty with this kind of study is that there are very few large Forbush decreases when satellite cloud data are available, so results tend to be dominated by a single event. This possibility is increased because the authors reduce the set of events to those common to all the available satellite cloud datasets used. The cloud response peaks 7 days after the GCR decrease, which is a not an expected delay. In the study of Kristjánsson et al. (2008), the number of samples was increased by taking a lower threshold cosmic ray decrease to define events. In addition to cloud cover, they studied cloud droplet size, cloud water content and cloud optical depth—all from the MODIS satellite dataset. The satellite data were divided into six regions, all of which showed a negative correlation between GCR fluxes and both cloud droplet size and cloud optical depth, but only one of which was statistically significant. For cloud cover and liquid water path, the correlations with GCR were weaker, with large variations in the behaviours between the different regions. Correlation coefficients were improved when only the six largest events were included, with 16 of the set of 24 correlations (four for each geographical region).
showing the sense expected by the hypothesis that GCRs help to seed clouds, but unlike the Svensmark et al. (2009) survey, correlations were not improved by introducing a lag. Sloan & Wolfendale (2008) also found that the correlations had no statistical significance, and Erlykin et al. (2009b) divided clouds into two major types and still found that the data gave no support to the contention that there is a large-scale causative cosmic ray cloud correlation, even for the parts of the world where correlations have previously been reported.

Erlykin et al. (2009a) used the method of Lockwood & Fröhlich (2007) to derive the solar cycle means and to study the well-known 22-year oscillation of GCRs. They detected a very similar residual oscillation in GMAST data after subtracting an assumed exponential rise, attributable to GHGs (but not allowing for volcanoes nor the ENSO oscillation). The correlation of this residual was higher for GCRs than for sunspots although the sunspots were in phase; the GCRs lagged 2 years behind which would be evidence that irradiance phenomena like TSI would have to be the cause rather than the GCRs (were the correlation indeed significant). In contrast, Harrison (2008) detected the 1.68-year oscillation, known to be in GCR fluxes but largely absent in TSI data (Rouillard & Lockwood 2004), in long sequences of ground-based cloud measurements. This suggests a GCR-induced effect on cloud rather than an irradiance effect.

In this area, indirect suggestive evidence is regularly reported. For example, Dengel et al. (2009) have recently published a small but significant correlation between cosmic ray fluxes and tree growth rates in the UK. On its own, one such correlation means little for climate studies, but in the context that most paleoclimate evidence for solar influences of climate involves reported correlations between cosmogenic isotopes and proxies for local and regional rainfall rates (see review by Lockwood & Fröhlich 2007), they may take on additional significance.

8. The current solar minimum in context and the future

Lockwood et al. (2009) have recently re-evaluated the derivation of open solar flux and mean solar wind speed from geomagnetic activity data, using a better correction for the processing of the interplanetary field by longitudinal solar wind structure between the surface of the Sun and the Earth. The results are very similar to those of Rouillard et al. (2007), who used an averaging procedure to effectively make the same correction. The implications of the derived changes in open solar flux have been reviewed by Schüssler & Baumann (2006) and Lockwood & Owens (2009). From the plot shown in figure 4, Lockwood et al. (2009) argue that the current low solar cycle minimum is part of the fall from the grand solar maximum (Usoskin et al. 2007) that has persisted during the space age. From the linear extrapolations (over short intervals) shown in figure 4, these authors predicted that the Sun will fall out of the maximum (defined by the mean level exceeded in 1920) within the interval 2011–2027. This agrees with the independent predictions for a consistent threshold by Abreu et al. (2008), based on the distribution of durations of past grand solar maxima in cosmogenic isotope data (the current grand maximum has already lasted for an unusually long time).

Abreu et al. (2008) employed the composite of three different cosmogenic isotope records compiled by Steinhilber et al. (2008), who also used modern neutron monitor data and Monte-Carlo simulations of the production of
cosmogenic isotopes to allow the composite to cover recent solar cycles. Importantly, the composite is generated using both the $^{10}\text{Be}$ and $^{14}\text{C}$ cosmogenic isotopes because the deposition of either is subject to climate effects (as, for example, discussed by Aldahan et al. 2008), but because the deposition of the two isotopes into their terrestrial reservoirs is so radically different, the combination is less likely to contain an effect of climate change. Steinhilber et al. (2008) also removed the effects of changes in the geomagnetic field. This composite is shown in figure 5, which plots the derived heliospheric modulation potential $\phi$ over the past 9000 years (we can think of $\phi$ as the effective retarding potential needed to turn the local interstellar GCR spectrum into that found in near-Earth interplanetary space). The 1920 level is $\phi = 600 \text{MV}$ (horizontal dashed line), and vertical dashed lines mark the ends of the maxima when $\phi$ fell back below the 600 MV threshold. In this record, there are 24 such events.

Predicting the behaviour of the solar dynamo is not currently possible (de Jager 2008); however, we can look at the range of past behaviours. A superposed-epoch composite of the variations in $\phi$ is shown in figure 6, where time $t = 0$ is when $\phi$ falls through the 600 MV level (the vertical dashed lines in figure 5). From figure 5, we can see that during the MM, $\phi$ is near 150 MV and there are a number of minima where $\phi$ reaches similar values and several where $\phi$ is lower. We here adopt $\phi = 175 \text{ MV}$ as a threshold, which defines a grand solar minimum that encompasses the MM. In figure 6, two traces show a fall to below this level within $t = 40$ yr. This means that past experience tells us that there is an 8 per cent chance that within 40 years from the predicted end of the current grand solar maximum (i.e. by 2060), the Sun will have returned to MM conditions. In contrast, there is also an 8 per cent chance that $\phi$ will only have fallen to 550 MV
Figure 5. Composite variation of the solar modulation parameter $\phi$ derived from cosmogenic isotopes, with allowance for the geomagnetic field variation, by Steinhilber et al. (2008). The plot shows 25-year averages. The vertical dashed lines mark the ends of intervals of high solar activity, defined by the $\phi = 600$ MV level (the horizontal dashed line). The horizontal dot-dashed line is the 175 MV level below which $\phi$ drops during a number of grand solar minima, including the MM.

Figure 6. Superposed-epoch study of the descent of the heliospheric modulation parameter $\phi$ at the end of a grand solar maximum. For each line $t = 0$ is the time when $\phi$ falls below 600 MV in the composite reconstruction by Steinhilber et al. (2008) (i.e. the times marked by vertical dashed lines in figure 5). The 24 lines are the variations for the 24 available cases in the last 9000 years. The horizontal dot-dashed line is the 175 MV level as given in figure 5. The thick dashed line shows the data for the past 100 years, showing the progress of the current grand solar maximum.

by this time. The probability of at least one grand minimum ($\phi < 175$ MV) having occurred within an interval of duration $t$ of the end of a grand maximum is shown in figure 7 for $t$ between 1000 and +1000 years. The fact that the distribution is asymmetric is not surprising because $t = 0$ is at the end (rather than the centre)
of the grand maximum, and there is also a tendency for the declines to be more rapid than the rises. Hence, although there is a chance that the current solar minimum heralds a rapid return to MM conditions in the next century, a smaller fall is actually more likely and MM-like conditions are most likely to be 100–200 years into the future.

Lean & Rind (2009) used extrapolated input forcings with a multivariate fit analysis to estimate that the solar decline may give a plateau in GMAST between 2009 and 2014. The predicted global maps show that the change will depend considerably on location.

9. A climate response to the decline in solar activity?

The recent solar minimum shows that the trends to lower TSI and greater cosmic ray fluxes identified by Lockwood & Fröhlich (2007) have continued and these would be, if anything, a cause of lower temperatures on Earth. It is not surprising, therefore, that the GMAST record has been scrutinized for a matching decline.

As shown in the bottom panel of figure 1, the NASA/GISS data show a continued rise in GMAST, whereas the HadCRUT3v data show a plateau after about 2000. This highlights the difficulty in combining the available data into a single homogeneous series. These datasets do not use the same input data, nor do they handle regions of missing data in the same way (in particular, GISS interpolates into areas of missing data, whereas HadCRUT3v does not). In addition, the intercalibration of the measurements must constantly be evaluated. A good example of the potential effects is the downward step in the GMAST of 0.3°C in the middle of the twentieth century, recently revealed by Thompson et al. (2008) and attributed to sea surface temperature measurement...
biases. As measurement techniques continue to evolve, such metrology issues remain important and, for example, the effect of the use of drifting buoy measurements on the recent data has to be evaluated (although the effect is expected to be small).

**Lockwood (2008)** carried out a simple multivariate fit analysis that fits the plateau in the HadCRUT3v GMAST data reasonably well. In this fit, the plateau arises mainly because of the lack of El-Niño events since the major 1998 event, with a small contribution from the solar decline. It must be stressed that inter-correlations between the input data sequences (for example, between TSI and the El-Niño index or the volcanic optical depth index and the El-Niño index) mean that this kind of analysis cannot prove that these are the contributions to GMAST; the best evidence requires full detection–attribution techniques with global climate models (see review by **Ingram 2006**). However, it does show that this feature could be readily explained as the combined effect of known influences. **Lockwood (2008)** passed each input variation through a low-pass filter and allowed each forcing to have a different response time, thus ensuring that the lag and smoothing were consistent. The lag is an important issue because, as shown by **Lockwood & Fröhlich (2008)**, solar inputs peaked between 1985 and 1987 and any plateau effect is after the major El-Niño year, 1998.

But, is the plateau in GMAST in the HadCRUT3v dataset unusual? This has recently been reviewed by **Easterling & Wehner (2009)** and **Knight et al. (in press)**. A simple evaluation of this question is given in figure 8. In this plot, running means have been applied to all the monthly HadCRUT3v GMAST

---

**Figure 8. Distributions of the probabilities that the last \( \tau/2 \) years show a temperature trend \( d\Delta T/\Delta t \) (defined positive for a rise), for various values of \( \tau \) ranging from 6 to 26 years in steps of 4 years. In each case, the monthly mean GMAST anomaly data are used from the HadCRUT3v dataset and a smoothed variation taken using the running means over intervals \( \tau \) long. The distribution of the deviation of monthly means from the running means is then used to calculate the probability of the deviation of the monthly means over the past \( \tau/2 \) years from linear rises that were varied between \(-10\) K and \(+8\) K per century.**
anomaly data, $\Delta T_S$, using averaging intervals $\tau$ long. The deviations of monthly values from this smoothed running mean are then evaluated and averaged up into intervals $\tau/2$ long. In this way, the observed probability distribution function (pdf) of these deviations from the smoothed average is computed. For the interval $\tau/2$ long at the end of the sequence (for which no running means are available), the mean deviations are taken from linear variations of slopes $d\Delta T_S/dt$ that varied between $-0.1$ and $+0.08$ kyr$^{-1}$, and the probability of each of the resulting mean deviations can then be read off from the pdf for that $\tau$. This procedure was repeated for $\tau$ from 6 to 26 years in steps of 4 years. The probabilities obtained for each $\tau$ are shown in figure 8 as a function of the drift over the last ($\tau/2$) years, $\Delta T_S/dt$. These give the probability that the recent slope has a certain value, based on the past behaviour. The plot shows that if we take $\tau = 6$ yr (i.e. we assume that we can define a trend in just $\tau/2 = 3$ yr), the most probable value for $\Delta T_S/dt$ is approximately $-0.02$ kyr$^{-1}$ (but there is a considerable chance that the trend could be upward). In other words, the temperature trend in recent years has probably been downward if we consider 3 years to be sufficient to define a trend. However, considering the frequency of events such as volcanoes and El Niño (and the solar cycle), $\tau/2 = 3$ yr is certainly not adequate to define a trend. If we increase $\tau$ to 10 years, figure 6 shows that the most likely trend is zero, and if we take $\tau = 14$ yr, the most likely trend is $+0.01$ kyr$^{-1}$. Increasing $\tau$ further makes the optimum trend approach $+0.02$ kyr$^{-1}$. In other words, one has to take a short-term view (less than 5 years) to see a recent downward trend, and a longer term view (more than 10 years) shows that any recent plateau is entirely consistent with the level of fluctuations seen previously about the upward trend.

Scafetta & West (2007) and Scafetta (2009) have proposed that the use of a single response time is not adequate and the climate system has a relative rapid response of approximately 1 year and second of about a decade. Using a combination of the two, they then deduce from a multivariate fit that 65 per cent of the GMAST rise can be attributed to TSI change. Section 10 considers the energetic implications of this. Neither Lockwood (2008) nor Scafetta (2009) quotes the statistical significance of their best multivariate fits, but Lockwood (2008) did compare the 2$\sigma$ deviations from the best fit for GHG and solar forcings. There are a number of complications with this kind of fit; for example, unknown inter-correlations between the inputs can influence the results. In evaluating the significance of any correlation derived one must allow for the degrees of freedom in the fit. Lockwood (2008) used seven fit parameters (a weighting factor for each of four inputs plus a lag for each except for the GHG forcing as it was taken to vary linearly with time). The formulation of Scafetta (2009) would, to be completely general, require 16 fit parameters (two lags for each of four inputs, which should all be treated in the same way, with a weighting function for each input/lag pairing). The significance of any such multivariate fit is dramatically reduced (especially when the effect of the autocorrelation of each sequence in reducing the effective number of independent samples is also considered). Further comments have been made by Benestad & Schmidt (2009). One key factor to note is the importance to the analysis of Scafetta (2009) of using the ACRIM TSI composite, on the grounds of the analysis of Scafetta & Willson (2009), but as discussed in §5, the justification for this is based on the application of an entirely inappropriate TSI reconstruction. The reason that this matters is that the PMOD composite shows that the mean TSI has fallen since 1985 (Lockwood & Fröhlich 2007), so that even
the decadal scale lag proposed by Scafetta (2009) cannot explain the fact that temperatures rose until 1998 unless the ACRIM composite is used. The ACRIM TSI composite is the only one that shows a long-term peak in 1992 (see figure 4 and Lockwood & Fröhlich 2008), which would allow the long time-scale response proposed to match the apparent plateau in the HadCRUT3v GMAST data.

Lastly remember also that this plateau is not present in the GISS data, which extrapolates available data into areas where data are missing (whereas HadCRUT3v does not). This means it is possible that HadCRUT3v underestimates the rise because of a lack of data where the recent warming has been greatest (e.g. the Arctic). However, the GISS extrapolation cannot be used as evidence that this is the case.

10. Concluding remarks

If we return to the energy balance equation (2.2), rearranging and differentiating it for the small changes observed, we get

$$4\sigma T_S^3 \Delta T_S + \Delta \gamma = \Delta I_{TS} \frac{(1 - A)}{4} - I_{TS} \frac{\Delta A}{4} + \Delta G.$$ (10.1)

From the GMAST record, we know that we need to explain $\Delta T_S$ of $+0.8$ K since pre-industrial times, around a mean value of $287.4$ K. This makes the left-hand side of equation (10.1) equal to $4.30 + 0.85 = 5.15$ W m$^{-2}$ (adopting $0.85$ W m$^{-2}$ for the current value of $\gamma$ and assuming that Earth was in radiative equilibrium in pre-industrial times). In §2, we discussed how observed GHG changes give a total radiative forcing (contribution to $\Delta G$) of $2.45$ W m$^{-2}$. Thus feedback loops, through their net contribution to the $\Delta A$ and $\Delta G$ terms, are required to supply a further $2.46$ W m$^{-2}$. In other words, the feedback must essentially double the GHG forcing if they are the cause of the GMAST rise. On the other hand, our best estimate of the first term on the right-hand side of equation (10.1) is $0.23$ W m$^{-2}$ ($\Delta I_{TS} = 1.3$ W m$^{-2}$). If the analysis of Scafetta (2009) were correct and 65 per cent of the observed warming were due to solar effects, then the first term on the right-hand side of equation (10.1) plus the feedback would need to supply $0.65 \times 5.15 = 3.35$ W m$^{-2}$. In this case, the feedback must supply $3.35 - 0.23 = 3.12$ W m$^{-2}$, which means that they need to explain an amplification of the solar input by a factor of 13.5. These feedbacks are prescribed in the fit weightings of energy balance models, whereas it is simulated as an emergent property in global coupled ocean–atmosphere models. Although these can reproduce an amplification factor of 3, no credible model has generated a factor of over 13. Furthermore, the remaining 35 per cent ($1.80$ W m$^{-2}$) is $0.65$ W m$^{-2}$ smaller than the known effect of observed rises in GHG concentrations. Thus advocates of the huge solar amplification (positive feedback) factor must also explain why the feedback to greenhouse forcing is at the same time negative. The issue is not ‘can the GMAST curve be fitted with combinations of solar variations’—with enough free variables the answer will be ‘yes’ (but such fits would have very low or no statistical significance): the challenge for attempts to show such a phenomenon could be real is to give credible explanations of feedback amplifications of more than 13 within the constraints set by observations and their uncertainties (and yet still give negative feedbacks to GHG forcings).
Just how poor and ill-informed some of the debate appearing on the Internet can become is illustrated by recurrent reports that global temperature rise is associated with changes in the corpuscular emissions of the Sun. The total energy input from the thermal solar wind plus suprathermal solar particles into the atmosphere and inner magnetosphere (some of the latter may be deposited in the upper atmosphere at a later time) is of the order of $10^{13}$ W or, per unit surface area of the Earth, 0.02 W m$^{-2}$. Even if we take the extreme case that this input was entirely absent during the MM (known not to be the case), we would require an amplification by a factor exceeding 250. Furthermore, this very little energy is deposited in the upper atmosphere (the thermosphere) and there is no known viable mechanism in the published literature that will allow it to influence the global troposphere, let alone with this huge amplification factor. It is true that the history of solar–terrestrial physics shows that one cannot use the absence of a known mechanism as more than an indication: Lord Kelvin famously dismissed the growing evidence for solar influence on the geomagnetic field as ‘mere coincidence’, using an argument based on magnitudes (Kelvin 1893). The argument turned out to be wrong because it did not allow for the existence of the solar wind (which was not suggested until 1901 by George Fitzgerald). However, that situation is not at all analogous to the present situation concerning climate change. Lord Kelvin was not proposing an alternative explanation and he was quite right to point out that chance agreements do occur in datasets of limited duration (but wrong to dismiss the possibility that it was real). In the case of climate change, there is no doubt that global mean temperatures have risen, so that the effect is known to be real. Furthermore, there is a viable explanation of that effect, given that the amplification of radiative forcing by trace GHG increases by a factor of about 2 is reproduced by global coupled ocean–atmosphere models. What is alarming is that in the face of this strong scientific evidence, some Internet sources with otherwise good reputations for accurate reporting can still give credence to ideas that are of no scientific merit. These are then readily relayed by other irresponsible parts of the media, and the public gain a fully incorrect impression of the status of the scientific debate.

The direct influence of cosmic rays on cloud albedo is much harder to put in context. If it has operated alongside GHGs, but there were no climate feedbacks, its effect on the term containing $\Delta G$ must have exceeded that of the term containing $\Delta A$ by the total 2.46 W m$^{-2}$ attributed to feedbacks. To argue that it replaces the GHG forcing requires that one find major errors in the calculation of radiative forcing or errors in the experimental data on the rise of GHG concentrations: neither is a realistic possibility. What is certain is that the uncertainties and lack of homogeneity in long datasets is a real problem for the evaluation of any such effect (i.e. for quantifying its contribution or finding if it exists at all).

It is important not to make the mistake made by Lord Kelvin and argue that there can be no influence of solar variability on climate: indeed, its study is of scientific interest and may well further our understanding of climate behaviour. However, the popular idea (at least on the Internet and in some parts of the media) that solar changes are some kind of alternative to GHG forcing in explaining the rise in surface temperatures has no credibility with almost all climate scientists.

Proc. R. Soc. A
The author is grateful to a great many scientists for valuable discussions and, in particular, to Giles Harrison and William Ingram and the referees for their comments on this manuscript. He also thanks Friedhelm Steinhilber of EAWAG for generating the cosmogenic isotope composite and for valuable discussions. The author also thanks the following for compiling and providing data: the Space Physics Data Facility at NASA/Goddard Space Flight Centre for the OMNI2 interplanetary dataset; the NASA/Goddard Institute for Space Studies for the GISS GMAST data; the UK Met. Office and the Climate Research Unit, Climatic Research Unit, University of East Anglia for the HadCRUT3v GMAST dataset; the Solar Influences Data Analysis Center of the Royal Observatory of Belgium for the sunspot number data; University of Oulu/Sodankylä Geophysical Observatory for the Oulu neutron monitor data; the Radiation Centre of World Physikalisch-Meteorologisches Observatorium, Davos for the PMOD TSI composite, the Active Cavity Radiometer Irradiance Monitor Science Team, Columbia University for the ACRIM TSI composite and the Royal Meteorological Institute of Belgium for the IRMB TSI composite.

References


