WHAT CAUSES CYCLE-RELATED GLOBAL SOLAR CHANGES?

J. R. KUHN
Department Physics-Astronomy, Michigan State University, E. Lansing, MI 48824, USA

ABSTRACT  Both helioseismic and photometric solar observations imply that the outer “skin” of the sun changes with the solar cycle. The changes are obviously well correlated with magnetic activity – but are they caused by, for example, incidental perturbations in heat flow around surface magnetic features, and/or can the photospheric changes be interpreted to tell us something about how deeper parts of the convection zone change with the cycle? This discussion will illustrate how precise photometry (spatially resolved and unresolved), helioseismology, and numerical simulation data are critical ingredients in the solution to this problem.

INTRODUCTION

Much of this meeting has been devoted to understanding the solar properties that are responsible for producing the observable characteristics of solar p-modes. We have been reminded here that the physics (i.e. restoring forces) and excitation mechanism (i.e. convective turbulence) are convincingly reconciled with the observed properties of the solar p-mode spectrum. We’re now learning about solar structure at the level of $10^{-3}$ departures from the spherical mean stellar model we call the standard model. Probably no one here would argue that the basic dynamics of the observed p-mode spectrum is not well understood.

Of course there is at least one other “mode” for which the theoretical situation is much different. The 11(22) year solar activity(magnetic) cycle is a solar “mode” for which we cannot write down a Sturm-Liouville equation. Furthermore, while many would argue that the basic physics here is the $\alpha - \omega$ dynamo mechanism, this is by no means universally accepted. I have in mind the fact that the internal rotation gradient (now measured helioseismically) isn’t what the dynamo theorists expected. Other’s have described the solar cycle in very different terms, e.g. “helical waves” (Bracewell 1988 ), an “extended solar cycle” (Wilson 1988), or perhaps some type of “torsional oscillation” is responsible for what we see as the solar activity cycle. It is fair to say that the physics of this mode is not well understood.

A large body of helioseismic data convincingly show that the solar oscillations are sensitive to the 11 year mode. Two other global solar properties that vary with the cycle are the solar luminosity and surface magnetic field. Can we make a simple physical connection between these observables and our restricted understanding of the 11 year mode? Note that a realistic picture must
involve the magnetic field – the fractional variation in the mean solar poloidal (or toroidal) field is unity whereas, as described below, the fractional variation in luminosity and mode frequencies is three and four orders of magnitude smaller. The small magnitude of these changes probably justifies the assumption that the oscillation and brightness changes are not dynamically important to the mode, i.e. that brightness and oscillation data may be used as diagnostics without concern for their “backreaction” on the underlying long period oscillation.

SOLAR FLUX AND LUMINOSITY CHANGES

The ACRIM and Nimbus/ERB experiments proved that the solar irradiance is changing on solar cycle timescales and shorter. The timeseries of irradiance data from ACRIM (Willson and Hudson 1988) is shown in Figure 1. The vertical scale is normalized to the mean 1986 (solar minimum) flux. It is clear from these data that the scale for the flux variations on both short and long timescales is $10^{-3}$ of the mean irradiance. This is an interesting amplitude since we know that the fractional area of the sun covered by sunspots near solar maximum (in units of solar “hemisphere’s”) is comparable, i.e. a few thousand’s. Since the bolometric contrast of sunspots compared to the quiet photosphere is close to unity the obvious conclusion seems to be that sunspots effectively block outflowing radiation from the solar interior to diminish the total flux observed at the earth (cf Hudson 1988). Another demonstration of the effect is to correlate the measured surface magnetic fields (which are indirectly and imperfectly a measure of sunspot coverage) with the irradiance data. Figure 2 shows the scatter diagram that results after high-pass filtering the irradiance data (by subtracting a 60 day running mean) and daily mean magnetic field measurements from Kitt Peak (Harvey 1991). There is a quite significant negative correlation between magnetic field and flux on timescales shorter than two months, i.e. sunspots make the sun appear darker.

That the sun is darker when the mean surface field is larger (on short timescales) doesn’t imply that the solar luminosity has changed. We already know that photospheric magnetic fields result in anisotropic radiation fields. The limb-distance dependence of the facular contrast function is proof enough of this. It is clear that the vertical magnetic fields associated with sunspots must decrease the radiation flux perpendicular to the photosphere. Is some or all of this energy deficit redirected tangent to the photosphere? One way to look for such an effect in the no-spatial-resolution, but high precision solar irradiance data is to look at the autocorrelation signal. Since the sun rotates with an average period of about 27 days, any flux deficit which occurs as a large spot passes near the center of the disk should be preceded and followed by a flux excess as the magnetic region passes near the east and west limbs about 1/4 of a revolution earlier and later. Thus this anticorrelated signal should appear as a negative dip in the autocorrelation at 1/4 of the solar magnetic feature rotation period. Figure 3 shows the autocorrelation of the high-pass filtered ACRIM data for a one year period near solar maximum and near solar minimum. In both cases there is a conspicuous negative peak at a lag of approximately 7 days. Notice that the negative peak has similar relative amplitude (compared to the variance of the data) when the relative facular and sunspot magnetic
CYCLE-RELATED CHANGES

FIGURE I  Normalized ACRIM Irradiance Measurements

FIGURE II  Short Period Irradiance-Magnetic Field Correlation
FIGURE III  ACRIM Autocorrelation (at Solar Min. and Max.)

correlations are different. Photospheric magnetic features redistribute radiant energy away from the normal direction.

This correlation between magnetic fields and flux has the opposite behaviour if the low frequency part of these timeseries is retained. Figure 4 shows a scatter plot of the irradiance and mean magnetic field without temporal filtering. It is clear that the magnetic field is now correlated with the irradiance. On the solar cycle timescale the sun appears brighter when it has a larger mean magnetic field near solar maximum. Is this a consequence of the changing latitudinal distribution of surface magnetic fields — and therefore simply due to a mean redistribution of radiation into the plane of the ecliptic, or is there a real change in the solar luminosity during the solar cycle?

Of course to answer this question we must have high precision solar photometric observations with spatial resolution and enough of a timeseries to average over the fluctuations in the brightness due to solar magnetic activity variations. The Mt. Wilson limb photometry data can tell us something about how the latitudinal surface brightness of the sun changes during the solar cycle. I argued (Kuhn 1991) that these precise differential photometric observations implied that the solar luminosity was changing during the solar cycle if the non-facular surface brightness component that was detected at the limb radiated approximately isotropically. More recent data (Lin 1992) now confirm that this flux excess is more isotropic than the facular radiation pattern. Lin used a CCD to measure the full-disk surface brightness profile and found that the limb brightness excess was indeed visible at disk center. Figure 5 shows the solar luminosity variation implied by the limb photometry data. The irradiance data from the ERB experiment and the sunspot number are also plotted here. It is evident that most of the irradiance variation between 1983 and 1990 is actually due to fractional changes in the solar luminosity at the level of $10^{-3}$. 
FIGURE IV  Long Term Irradiance Correlation

FIGURE V    Long Term Solar Luminosity Variations

© Astronomical Society of the Pacific • Provided by the NASA Astrophysics Data System
FREQUENCY SHIFTS

Helioseismic observations show time dependent frequency changes in both the mean observed eigenfrequencies and in the amplitude of the m-dependent frequency splittings. Consider the usual empirical expansion for observed modal frequencies $\delta \nu_{nlm} = \sum_k b_{nlk} P_k(-m/l)$. The odd coefficients are sensitive to the internal solar rotation and the even coefficients $b_{nl2s}$ measure the axisymmetric solar structure. Notice that $b_{nl0}$ (as one of the even coefficients) quantifies the mean frequency shift (with time) in a mode multiplet. The shift coefficients have magnitudes of 100’s nHz but the typical measurement noise per multiplet is considerably larger. Thus, most of the data displayed below have been averaged over the 5 minute band in $n$ and $l$ indices. Figure 7 (below) is based on data collected from many sources which are described in Kuhn (1991). The most recent splitting data plotted in that figure comes from a preprint from Ronan and Labonte (1992). The scale of the changes in these coefficients is also 100’s of nHz and, since the measurements have been collected from different observations using different mode sets in the average, some care must be exercised in interpreting the variations. Yet, it is clear that the mean multiplet frequencies do change with the solar cycle along with the splitting coefficients. Clearly the magnitude of the frequency changes and the phase at minimum, relative to the time of solar maximum, vary from $b_0$ up to $b_4$ – the highest order term for which we have a long timebase of data. Note, in particular, that the magnitude and phase of mode frequency shifts during the cycle are quite comparable to the splitting coefficient changes.

Libbrecht and Woodard (1990) were the first to show convincing data that had a significant frequency dependence to the frequency difference between modes observed at two different times. They found that the temporal shift peaked for modes with eigenfrequencies near 4.3mHz and dropped off rapidly at higher frequencies and more slowly to lower frequencies.

All of the data described above measure eigenfrequency changes on timescales longer than a few months. The irradiance data show significant solar changes on timescales as short as a day (and less), shouldn’t we expect similar frequency changes? Figure 6 shows the mean frequency shift between 1986 and 1989 computed from BBSO data that was analyzed in roughly 1 month time segments (Woodard et al. 1991). The mean KPNO magnetic field during each of these periods is also plotted on this figure for comparison. It is evident that frequency changes are correlated with the mean KPNO-measured magnetic field on timescales much shorter than the solar cycle, as they are correlated with solar cycle changes in the magnetic field.

INTERPRETING THE HELIOSEISMIC DATA

Several authors have tried to reproduce the amplitude and form of the frequency dependent shift (with various degrees of success) but all would agree that the outer part of the sun is responsible for the shift. A simple forward calculation shows that a region within about 200 km of the photosphere causes the frequency changes (Kuhn 1990). Shibahashi (1991) worked out asymptotic inversion formulae to determine the perturbation depth dependence, and reached
the same general conclusion. Since the higher frequency modes have their energy density strongly weighted toward the surface the increasing fractional eigenfrequency perturbation requires a near-surface change in the acoustic sound speed or density structure.

The frequency dependence, and in particular the rapid drop at 4.3 mHz, might be due to a combination of magnetic fields and entropy perturbations near the photosphere. Along these lines, Roberts and collaborators (cf. Evans and Roberts 1990) have built simple exact models, while Goldreich et al. (1992) used a perturbative calculation to reproduce the form of the frequency shifts. Unfortunately both models do not address the issue of the other solar cycle observables. In particular the proposed near-photospheric perturbations should be consistent with observed solar luminosity changes and constraints on globalscale shape and surface brightness changes in the photosphere. The problems of Reynolds stresses and the anisotropic radiation field near magnetic regions are also not easy to include in these models. As argued below, a local mixing length theory for convection also fails to capture the qualitative behavior observed in numerical convection experiments when corresponding perturbations have been induced.

Perturbations to the mean solar stratification may affect eigenfrequencies by changing the local sound speed and the “size” of the acoustic cavity. The frequency shift and splitting data could be a consequence of changes in the outer boundary conditions (the solar shape) or largescale changes in the sound speed associated with magnetic or thermal perturbations. A change in the stratification might be due to largescale variations in convection (Gough and Thompson 1988) or the local (non-kinematic) effect of magnetic fields. An observer’s approach to disentangling this problem starts with the observation that the solar shape changes are much smaller than the apparent entropy changes in the photosphere. For example, between 1983 and 1985 the fractional oblateness change
1993ASPC...42...27K

FIGURE VII  Comparison of Helioseismic and Photometric Frequency Shifts and Splittings

\((\delta r_2/R_\odot)\) of the sun was less than \(10^{-2}\) of the fractional change in the luminosity. Here the subscript 2 refers to the second order legendre coefficient of the limb shape. Similarly the magnitude of the oblateness is much smaller than the corresponding limb brightness term \((\delta r_2/R_\odot \leq 3 \times 10^{-3} T_2/T_\odot)\). Since the apparent changes in the boundary of the acoustic cavity are small, my approach was to use the limb brightness observations to estimate the change in the local sound speed (Kuhn 1988,1989), while ignoring shape variations.

A legendre expansion of the surface temperature distribution takes the form \(\delta T(\theta) = \Sigma s t_{2s} P_{2s}(\cos \theta)\). A simple asymptotic expression then relates the induced even splitting coefficients \((b_{2s})\) to the limb temperature coefficients \((t_{2s})\). In general \(b_{2s} = K_{2s}t_{2s}\) where the coefficient of proportionality \(K_{2s}\) depends on the radial dependence of the temperature/sound speed perturbation. For a simple model where the fractional soundspeed perturbation is independent of depth \(K_0, K_2, \ldots = 175, -140, 110\) nHz/K. Figure 7 shows a comparison of the helioseismic coefficients implied by the limb photometry observations, using these simple proportionality coefficients. It is perhaps surprising that the magnitude and phase of the splitting results are predicted so well by the photometry. Goode and Kuhn (1989) showed that if the centrifugal contribution and a radial dependence were allowed then the photometric and helioseismic results agree to within their observational errors.

Of course the BBSO data imply that the sun is perturbed near the surface, not with constant fractional amplitude with depth. Furthermore the calculations described were based on an average over the 5-minute band and so have ignored the frequency information contained in the splittings. Thus this picture is also incomplete but, as argued below, it does provide a basis for describing both flux and helioseismic data.

© Astronomical Society of the Pacific • Provided by the NASA Astrophysics Data System
INTERPRETING THE FLUX VARIATIONS

It is difficult to understand how the luminosity can vary over an 11 year cycle unless energy is "stored" somewhere in the solar convection zone (SCZ) during the solar cycle. Models which invoke periodic energy generation mechanisms in the core must have enormous non-linear modulation amplitudes in order to have measurable amplitudes outside of the radiative core – recall that the diffusion time through the core is thousands of years. The thermal "timescales" in the convection zone are strongly dependent on depth and complicated by the possibility of convective and radiative energy transport (Spruit 1988). The timescale for fluid from the base of the SCZ to appear at the photosphere is a few days or weeks, while the thermal timescale via any transport mechanism is much less than this within a few thousand kilometers of the photosphere. If we accept the philosophy that the changes in the sun that are responsible for the luminosity variations are not invisible to the solar acoustic modes, and vice-versa then we should ask how a perturbation in the outer skin of the sun can change the luminosity on such a long timescale. Again, since the natural timescales (radiative or convective) in the outer 200 km are much less than a few hours, any forcing function with a period of 11 years is strongly attenuated. This point is also illustrated by a simple calculation.

It has been argued that magnetic fields or photospheric flux tubes can provide low resistance channels for radiation to escape from below the photosphere, thus increasing the total radiated energy (luminosity). This might be the effect of faculae, for example, where the magnetic pressure inside the tube leads to a lower gas density and a longer photon free path. Faculae appear as "windows" into the photosphere. In a model where energy transport is by diffusion and where the free path is decreasing exponentially (like the reciprocal of the density in a solar atmosphere) it is not hard to show that such "holes" do not lead to a larger luminosity. Energy is much more effectively transported horizontally than vertically and so the excess flux in the tube ultimately comes from the surrounding layers, with very little change in the total luminosity from the full region.

As an example consider a 64×64 grid to describe a fluid with an exponentially declining conductivity with depth. In this calculation the scale-length of the conductivity was 4 grid lengths. At the center of the region is an "antiplug" – a region with twice the conductivity of the surrounding layers. The plug was several scaleheights from the top and bottom boundaries of the full region. The equilibrium temperature distribution was now computed assuming constant temperature boundary conditions at the top and bottom. While the flux directly above the antiplug was about twice the flux in the undisturbed region, the total luminosity from the full region was hardly affected by the perturbation. The change in the luminosity compared to the area integral of the flux through an undisturbed region of the same size as the plug was only 10^{-3}. The perturbation must extend to within a couple of scale-heights of the bottom of the simulation area in order to increase the total luminosity by more than 1 percent. The implication of these observations is that the diffusion length and timescales are such that it may be difficult to poke "holes" in the photosphere to increase the luminosity, unless the photospheric perturbations cause mean structural changes in the solar stratification at depth. Of course this is not a
realistic model of convective energy transport and may not describe the solar problem. For this, numerical experiments are needed.

LUMINOSITY CHANGES

So how can the solar luminosity be affected? Sunspot polarities and their cycle evolution seems to require the growth and evolution of a toroidal magnetic field somewhere in the solar interior. Where else can this field exist but at the base of the convection zone? We know that rotational shear can generate a toroidal field. Furthermore, helioseismic measurements (Schou 1991) of solar rotation place the region of greatest shear at the radiative-convection zone boundary. Let us assume that, by some mechanism, the "solar cycle" generates a toroidal flux "sheath" immediately below the SCZ.

As the field develops pressure equilibrium requires that the magnetic contribution be balanced by a diminished kinetic pressure, and thus a lower density at the same temperature in the field region. The radiative flux is proportional to $\Delta T^4/\kappa \rho H$ where $\Delta$ is the logarithmic temperature gradient, $\kappa$ is the opacity, $H$ is the pressure scale height, and $\rho$ is the density. Thus the reduced density increases the radiative flux through the field region to the SCZ above it. The increased radiative flux also increases the temperature of the fluid in the magnetic region so that the density decreases even more. This further increases the radiative flux from below until the density in the magnetic region immediately below the convection zone is convectively unstable. The result of this is that both excess entropy and magnetic flux should be advected to the surface above magnetic field regions. The lifetime $\tau$ for this instability can be estimated as $1/\tau = 16\sigma T^3/3\pi \kappa \rho ||dT/d\tau|| \delta \rho / \rho / l$ where $\sigma$ is the Stefan Boltzmann constant, $\delta \rho$ is the initial density perturbation due to the magnetic field, and $l$ is the vertical scale of the perturbation. For a sharp magnetic zone boundary the growth time $\tau$ can be much shorter than the solar cycle timescale even if the local magnetic field strength is only 10KG.

NUMERICAL TOOLS

If the solar luminosity "valve" is located deep below the photosphere then why isn't this perturbation well mixed by the SCZ and what produces the surface perturbation implied by the helioseismic data? It has been argued that convective flows must effectively couple most of the SCZ to any local perturbation. Spruit (1988) describes the effective "conductivity" of the SCZ as nearly-isotropic and "large." In this case the surface brightness changes are likely to be determined only by the surface perturbation, i.e. the magnetic valve must lie near the surface. A latitudinal variation in entropy perturbation at depth would then have an uncomfortably large amplitude to be measureable at the surface.

Stein and Nordlund (1989) can realistically simulate a piece of the outer SCZ, at least at the level of reproducing what we can see of the photospheric convection properties from earth-based observations. Figure 8 illustrates a slice of one of their deeper simulations. The simulation volume is 9000km deep and 12000km on a side. The darker shades show cooler regions and the arrows indicate particle velocities. The correlation length of, for example, the cold plume
CYCLE-RELATED CHANGES

FIGURE VIII  Solar Convection Exposed by Numerical Experiments

on the right is much longer than the classical mixing length. The correlated downflow extends over nearly the full simulation cube height, yet the density scale height is an order of magnitude smaller.

We performed some experiments with a smaller volume approximately 3100 km on a side. In the left half of this cube we increased the entropy of incoming fluid at the bottom by 1%, while on the opposite side we decreased the entropy by the same factor (Kuhn 1991). After waiting for about three rise times (90 minutes of solar time) we checked to see how the entropy perturbations were mixed toward the top of the volume. We were suprised to learn that the mean temperature difference along a horizontal surface in the left and right halves of the volume actually increased with height near the surface. Figure 9 also shows the difference in temperature between hot and cold regions. Not only is the horizontal convective mixing generally ineffective, but in the superadiabatic region near the photosphere (where the convective efficiency is low) the temperature difference between hot and cold regions is amplified. There is some horizontal mixing in the lower 2000 km, but as the atmosphere becomes transparent a larger temperature gradient is needed to carry the same convective flux. In a linear diffusion model the vertical temperature difference should be exponentially damped by horizontal mixing. From the observed decline in $dT$ with height near the bottom of the model one can estimate that the horizontal “conductivity” is at least 10 times smaller than the vertical conductivity. The transport properties of the SCZ are evidently quite anisotropic, as the image of convective flows in Figure 8 suggests.

The eigenfrequencies of acoustic modes that propagate in the mean density stratification defined by the hot and cold sides of the simulation box were also calculated. The “hot” modes had frequencies which were approximately 2% higher than modes in the “cold” side. The temperature difference profile of figure 9 shows a rapid decline and reversal in the convective overshoot region.
FIGURE IX  Temperature Variations in the Hot-Cold Convection Experiment

where the more vigorous convective flows of the hot side of the simulation lead to stronger penetration into the stable layers above. This causes the sharp change in the temperature difference profile. This behaviour may cause a rapid falloff in helioseismic frequency shifts at high frequencies, in qualitative agreement with the observations. A quantitative comparison will require higher spatial resolution and a larger simulation volume.

CONCLUSIONS

Occam's razor leads us to the assumption that the solar magnetic cycle controls changes in the SCZ which affect both helioseismic and luminosity observations. An interesting question is whether it is changes in the upper atmosphere due to the photospheric magnetic fields which drives these variations. The alternative presented here is that both magnetic flux and entropy fluctuations are advected from the base of the SCZ. The timescale for magnetic flux emmergence through the photosphere should also describe a timescale over which the solar luminosity may change, since advected fluid carrying flux also carries excess entropy. We have already noted that the p mode frequencies and splittings change on short timescales of a month or perhaps less. A definite result of this picture is that the solar luminosity should also vary on this scale. Of course this observation is difficult since the magnetic field also leads to an anisotropic radiation field – perhaps there is a statistical solution to the observational problem of detecting a time dependent luminosity from a single-direction flux measurement.
REFERENCES

Harvey, J. 1991, personal communication.