Irradiance Variability of the Sun

Claus Fröhlich

Physikalisch-Meteorologisches Observatorium Davos, World Radiation Center, CH-7260 Davos Dorf, Switzerland

Abstract: Direct measurements of the solar constant — the total irradiance at mean Sun-Earth distance — during the last ten years from satellite show variations over time scales from minutes to years and decades. At high frequencies the spectral power is determined by granulation, super- and mesogranulation. In the 5-minute range, moreover, it is dominated by power from the solar p-mode oscillations. Their power and frequencies change with time yielding information about changes in the convection zone. Towards periods of several hours the power is steadily increasing and may be partly due to solar gravity modes. The most important variance is in the range from days to several months and is related to the photospheric features of solar activity: decreasing the irradiance during the appearance of sunspots, and increasing it by faculae and the magnetic network. Long-term modulation by the 11-year activity cycle are observed conclusively with the irradiance being higher during solar maximum. All these variations can be explained — at least qualitatively — by their manifestation on the photosphere. For the long-term changes the simultaneous changes of the frequencies of solar p-mode oscillations suggest a more global origin of the variations. Indeed, it seems that the observed irradiance modulation is a true luminosity change with the magnetic cycle of the Sun.

1 Introduction

The irradiance from the Sun at the mean Sun-Earth distance, integrated over the energetically important wavelength range (hence total irradiance) is called "solar constant". Observations of the solar constant have a long history, starting with the measurements by the Smithsonian Institution from mountain stations at the turn of the century and including data from the beginning of space era. Many of the early measurements were inconclusive, mainly because of lack of sufficient radiometric precision, but also due to influences of the Earth's atmosphere for observations from ground and airplanes (for reviews see e.g. Fröhlich 1977, Hoyt 1979, Angione 1981, Fröhlich 1987). The first clear evidence of solar constant variability only appeared at the beginning of the 1980's with the data from radiometers on the Solar Maximum Mission (SMM) and on NIMBUS-7, proving that the Sun is indeed a "variable" star. The interest in total solar irradiance variability on all time scales has then very much increased and the modern data on total solar irradiance reveal a broad spectrum of variations (e.g. Hudson, 1988). Not only are the atmospheric physicists and climatologists concerned, because of possible effects on the Earth's energy balance, but also the solar physicists have become interested: the existence of global changes of the solar output had been doubted for a long time and their observation obviously leads to new ways to understand the Sun.

The measurements from satellites discussed in this paper have been performed by the sensors of the Earth Radiation Budget Experiment (ERB) of the NIMBUS-7 satellite in a near-polar orbit since November 16, 1978 (Hickey et al. 1983 and 1989) and by an active cavity radiometer on the Solar Maximum Mission Satellite in a 27° orbit since February 14, 1980 (Willson 1984). Reference is also made to the data from the ERBE experiment on NASA ERBS in a 57° orbit and the NOAA9 and 10 satellites in near-polar orbits starting in October 1984, January 1985 and October 1986 respectively (Mecherikunnel et al., 1988). In all
these experiments the solar irradiance is measured by electrically calibrated cavity radiometers. The time series of the three experiments are plotted in Fig.1 and illustrate the solar irradiance variability on all time scales. The difference in value between the three experiment is due to their absolute calibration which is accurate to “only” about ±0.2% and does not reflect the precision and stability of the instruments which is obviously much better.

Furthermore data from the IPHIR investigation (Fröhlich et al. 1990a) on the USSR mission PHOBOS to Mars and its satellite Phobos are used. With this experiment the solar spectral irradiance in three wavelength bands centered at 335, 500 and 865 nm was observed continuously during more than 5 months between July and December 1988 before the spacecraft was injected into the orbit around Mars.

2 Variations on Time Scales up to Months

At high frequencies the most important features of the variance are the solar p-mode oscillations manifested as brightness fluctuations (e.g. Woodard and Hudson 1983, Fröhlich et al. 1990a, Toutain 1990). These oscillations are eigen-modes of the Sun and the restoring force is pressure, hence p-modes which are classified by degree l, order n and tesseral order m. These eigen-modes yield a equi-distant discrete spectrum due to the well defined resonant cavity. The values of the mode frequencies $\nu_{l,n,m}$ have important diagnostic power and are used in helioseismology, which is applying seismic techniques for sounding the interior of the Sun, as it is done with waves from earthquakes on Earth. An example of the p-mode spectrum from the IPHIR investigation is shown in Fig.2. In the context of this paper the long-term changes of these low degree p-mode frequencies and the variations of their amplitudes are important. The former will be discussed together with the solar cycle variations where possibly changes in the upper boundary of the resonant cavity are changed, whereas the latter is important for this section as these changes reveal information about short to medium term global changes in the upper layers of the convection zone. These may influence the luminosity and give a clue to the underlying physical mechanisms.

The solar variance is best illustrated by the power spectral density as shown for the SMM/ACRIM data in Fig.3 for 1980 and '85. It covers the range from about 50 nHz to 86 $\mu$Hz, the Nyquist frequency
corresponding to the orbital period of SMM. At low frequencies the strength of solar activity changes the spectral density by up to a factor of ten (1980 compared to 1985) whereas the density at higher frequencies seems independent of solar activity. In this frequency range the power is mainly due to granulation, super- and mesogranulation, which is illustrated by the result of an analysis of simulated time series (Andersen 1990). The simulations are based on a compilation of direct observations of such features on the solar disk both in terms of contrast and behavior with time. In the range above 10 \( \mu \text{Hz} \) (11.57 \( \mu \text{Hz} \) corresponds to a period of 1 day) the behavior of the ACRIM spectrum is very similar to the simulated one; the level of power is, however, three to five times higher. At lower frequencies the difference becomes quite obvious; in this range it is due to the fact that the simulations do not include any effects from solar activity. The difference at high frequencies may be an indication of power from solar gravity oscillations. Searches for these modes have been performed, but the results are not yet conclusive (Kotov et al. 1984, Fröhlich and Delache 1984, Henning et al. 1988, Kroll et al. 1988, Fröhlich 1990, van der Raay 1990). Their detection and identification, however, would be very important for constraining the structure and composition of the solar core.

In the frequency range above 10 \( \mu \text{Hz} \) the power is mainly due to activity related features on the Sun. The time scales associated with solar magnetic active regions reveal extremely interesting variations of the total solar irradiance. Here we will discuss time scales of a few solar rotations (about 0.1 \( \mu \text{Hz} \)) down to two days (the 5.8 \( \mu \text{Hz} \) Nyquist frequency of daily sampling). The dominant feature in the time series for this range of periods is the "dip", a negative excursion of a few days' length and a depth ranging up to a few tenths of a percent of the normal total irradiance. These dips result from large sunspot groups rotating past the central meridian. Between the dips the total irradiance fluctuates on a wide range of time scales, as demonstrated by the broad-band character of the power spectrum. Prominent dips appeared during the first few months of data from the ACRIM instrument on board SMM, and Willson et al. (1981) described them in terms of the Projected Sunspot Index, the PSI function similar to the models of sunspot darkness noted earlier by Foukal and Vernazza (1979) and Hoyt and Eddy (1982). Hudson and Willson (1982) give the definition of this projected-area index as follows:
Fig. 3. Powerspectrum of the solar irradiance for 1980 and 1985 observed by SMM/ACRIM (left, Fröhlich et al. 1990b) and of the simulated irradiance at 500 nm (right, Andersen 1990) due granulation, super- and mesogranulation.

\[ PSI = \alpha \cdot \sum_i \frac{3\mu_i + 2}{2} \cdot \mu_i \cdot a_i \]

where \( \mu_i \) is the central angle of the sunspot group \( i \) and \( a_i \) its area. The factor \( \alpha \) takes into account the umbra/penumbra area ratio and the effective temperature of the sunspot photospheres, in the simplest possible way. This index, calculated from the synoptic data for sunspot areas and with textbook calculation of the value of \( \alpha \) (i.e., no adjustment of parameters whatsoever), removed about half of the variance of the total ACRIM time series in the active year 1980. This success was quite surprising in view of the grossly simplifying assumptions involved in the construction of \( PSI \). For the period 1980 to 1988 the \( PSI \) function is shown in Fig. 4 together with the irradiance data from ACRIM and ERB corrected for spots by this function (\( S + PSI \)). The success of the simple \( PSI \) explanation of sunspot effects immediately implied that a “bright ring” surrounding the sunspot did not directly re-radiate the blocked convective flux presumed to be welling up continuously through the solar interior. The attempts to improve generally on the naive fit of the \( PSI \) have generally failed, owing perhaps to the additional noise in the correlation introduced by random and systematic errors in the synoptic data (Schatten et al. 1985).

Willson (1982) called attention to the effects of faculae in modulating the irradiance and having the potential to compensate the sunspot deficit. More direct photometric approaches have yielded some insight into the relationship between the irradiance excesses due to faculae of an active region and the sunspot deficits. Given the almost simultaneous existence of dark spots and bright plage in active regions, it immediately became interesting to ask if the energy excess and deficit actually balanced. If so, this would imply some mechanism for local storage via the sub-photospheric magnetic field structure. Indeed, at the simplest level both the Ca proxy measure of facular brightness (Hirayama et al. 1984, Chapman et al. 1986) and direct facular photometry (Chapman 1984, Lawrence et al. 1985) show a rough balance (see Chapman, 1987, for a complete review).

Multi-variate spectral analysis is a powerful tool for the investigation of multiple influences on a time series, as for example the quasi-independent spot and facular contributions discussed above. Fröhlich and Pap (1989) have applied this technique to the ACRIM data, both in 1980 and after the repair mission in 1984/85. The results are shown in Fig. 5. They find for the former period that more than 90 percent of the variance of the ACRIM data can be explained by the effects of sunspots and magnetic elements, the latter including both active-region and network facular elements (represented by the equivalent width of the He
Fig. 4. The influence of sunspots on the irradiance during 1980-88. Top panel: the downward plotted \( \Psi \) function and bottom panel: the \( \Psi \) corrected irradiance \((S + \Psi)\) of ACRIM and ERB with the solid lines indicating a 81-day running mean (from Fröhlich et al. 1990b).

1083 nm line, the HeI index). The analysis reveals the presence of power spectral peaks not explained by spots, faculae or magnetic network near 27 and 9 days, the latter suggestive of heretofore unknown processes within the convection zone.

The long stretch of the IPHIR data allows for the first time to study in detail the time variation of the amplitudes of the p-mode oscillations (Toutain 1990, Fröhlich and Toutain 1990). In Fig.6 the total power as sum of the individual power of 18 lines with \( l = 0 \ldots 2 \) and \( n = 18 \ldots 23 \) are shown as time series and estimates of the spectral power density. Most of the variation comes possibly from stochastic excitation and damping of the modes by convection. It is, however, interesting to note that the power is varying by nearly a factor three during the 160 days observations, although it is the sum of 18 individually excited and damped modes. Moreover, the estimated life time of the modes is of the order of a few days and their are periods of low or high power lasting for 20 to 30 days. Some of the variation could also be due to modulation by the irradiance itself, but the comparison with the irradiance variations shows in the time series no apparent correlation. For some periods a positive and for others a negative correlation may be observed. In the frequency domain the correlation is also not obvious. Between 0.5 and 1 \( \mu \text{Hz} \) similar features are observed in both spectra; the significance of this similarity, however, is not clear. It is interesting to note that the broad peak observed around 9 days as residual in 1980 (Fig.5) is also present in this spectrum with the same width (1.05-1.35 \( \mu \text{Hz} \)). The peak at 0.4 \( \mu \text{Hz} \) (25-33 day period) may be related to solar rotation, again it could be of the same origin as the unexplained 27-day peak in the 1980 and 1985 spectrum. The origin of the strong peak
Fig. 6. For 1980 (left) and 1984 (right) the power spectra of the SMM/ACRIM data and the results of the multispectral analysis are shown: the shaded areas indicate the part of the irradiance power explained by the power of active and passive spot areas (similar to PSI, which is the sum of both) and of the equivalent width of the 1083 nm He I line (from Fröhlich and Pap 1989).

around 0.2 µHz (50-96 day period) is not obvious, but again it is at the same place and has the same width as the lowest unexplained peak in the 1980 spectrum of Fig. 5. All three coincidences may be fortuitous, but it is highly suggestive that they may have the same origin which is related to the activity cycle. 1980 was right after the maximum of the cycle 21 and 1988 is close to the end of the ascending phase of cycle 22; a possible explanation of the better correlation with the 1980 data.

4 Variations on Time Scales of Months to Years

After smoothing the time series of the ERB and ACRIM radiometry (solid lines in Fig. 1 and 4) both data sets show common variations on time scales of 4-9 months, as well as a general slow downturn, which has reversed during 1986, after passage of the solar minimum.

Several ideas and models have been put forward to explain these variations. One approach is to account for the variations purely in terms of the effects that magnetic flux tubes seem to have on the radiation and convection in the relatively shallow photospheric layers that emit most of the Sun’s luminosity. Others have invoked deep seated changes in solar convection, perhaps involving even variations in the nuclear-burning core (e.g. Gough, 1989).

A relatively straightforward approach to both the 4-9 month and 11-year variations has been put forward by Foukal and Lean (1988) and Livingston et al. (1988). In these studies, it was shown that the residual irradiance variations remaining after a correction for sunspot blocking is made to the smoothed ACRIM or ERB data correlate well with indices of facular area such as provided by the He I index or the 10.7 cm flux. This is not surprising since facular area variations were previously shown to account for day-to-day variations in these residuals (Hudson and Wilson 1981, Foukal and Lean 1986).

The He I index and the 10.7 cm flux represent contributions from all bright magnetic elements on the disc, including the network. Thus, one may conclude that the 6-9 month variations are caused by the tendency of major complexes of activity to persist for about this number of solar rotations (Gaizauskas et al. 1983). This time scale in persistence of solar activity episodes has been documented before in studies of the He I index time series (Harvey 1984) and may be related to the persistence of certain active longitudes on the solar surface, noted already by Dodson and Hedeman (1970). No well-accepted explanation of this 6-9 month activity time scale (or of the active longitudes) exists as yet, although the recent ideas of Wolff (1984) in terms of Rossby waves and surface pattern of g-mode oscillations and their possible interference are interesting,
Fig. 6. Variation of total line power of solar p-modes as observed in brightness by the IPHIR investigation. Left: time series of total line power (heavy line) and of solar irradiance from SMM/ACRIM (thin line, the range from bottom to top is 1366.6 to 1368.8 Wm\(^{-2}\)). Both sets are running means over 8 days. Right: power spectral density of the total line power (heavy line) and the solar irradiance (thin line). The irradiance spectrum is shifted so that it is at the same level as the total line power at low frequencies.

and deserve more attention. Wolff and Hickey (1987) have also proposed that such oscillations may modulate heat flow to the photosphere and thus contribute to irradiance variation directly.

The finding that simulation of the downtrend between 1981-1986 requires use of a global index, as mentioned above, indicates that this downtrend is associated with some kind of a slow change in the solar atmosphere well modeled by e.g. the He I index or the 10.7 cm flux. The simplest explanation is a slow decrease in the emission from the bright magnetic network outside active regions. This identification cannot be considered proven until the changes in area (and possibly also intensity) of the network are measured in white light over a solar cycle. Fig. 7, lower panel, suggests that the whole solar cycle can possibly be modelled by these indices.

Kuhn et al. (1988) have reported observations of the limb brightness which can be used to explain the total irradiance variation of the solar cycle. The observations are broad-band, two-color photometric measurements of the brightness distribution in a narrow annulus 20 arcsec wide, just inside the solar limb. The solar limb flux observed as a function of latitude can be divided into a “facular” and “temperature” part based on the assumption that the “temperature” part is constant over the 4-month observing summer period and that the “facular” part shows up as intermittent bright regions. The component of excess brightness moves toward the equator between 1983 and 1985, and then reappears at relatively high latitudes again in 1987. The excess brightness responsible for \(dT\) is due to features which are not resolved by this observation, and it could be due to the bright network in and outside the active regions. The decreasing contribution of \(dT\) (including the facular component) can account for the total irradiance decrease between 1983 and 1987, and its increase in 1988 (Kuhn, 1989) as shown by Fig. 7.

Only the zero order term of the polynomial expansion of the latitudinal variation of the \(dT\) signal is responsible for the irradiance variation. In a separate paper Kuhn (1988) compared the expansion coefficients needed to explain the latitudinal variation of \(dT\), which can be interpreted as an asphericity. This breaks the spherical symmetry and can be observed as splitting and frequency shifts of the p-mode oscillations. He developed a model in which the underlying physical effects are parametrized by a local effective sound speed where \(\frac{c_s}{\omega}\) is directly related to the surface \(\frac{c_s}{\omega}\) observed. He is able with his model to explain consistently
Fig. 7. Comparison of the 81-day running mean of the observed irradiance from ACRIM with two different models: The \( PSI \) corrected ACRIM irradiance \((S + PSI)\) is shown as broken line, the irradiance residual calculated from the 10.7cm flux (the scaling is determined by linear regression of \( S + PSI \) and the 10.7cm flux) as solid line and the irradiance data deduced by Kuhn et al. (1989) from \( dT \) and the simultaneously measured facular contribution as diamonds (from Fröhlich et al. 1990b).

the p-mode splitting coefficients from the corresponding coefficient of the \( dT \). In a more recent paper Kuhn (1990) shows that the splitting data from the Big Bear Solar Observatory (Woodard 1990, Libbrecht 1990) are best fit by a shallow \( \delta r_\ell^2 \) perturbation, extending downward less than a few hundred kilometers below the photosphere. Also the observed centroid frequency shifts of the low order p-modes first reported by Woodard and Noyes (1985) fit well to this simple model as well as the frequency dependence of the \( l = 1 \) mode changes (Pallé 1990). These independent results strongly support the model that the observed \( dT \) is due to a global effect manifested by an asphericity in \( c^2 \) and in \( T_{\text{eff}} \) in a shallow layer below the surface which varies with the solar cycle. This also suggests that it is the solar luminosity and not only the observed irradiance that evolves during the solar cycle. Moreover, these results also imply that the fractional radius variations are much smaller than the fractional luminosity changes; thus \( \delta r_\ell / \delta L / L \lesssim 0.03 \) is very small.

How this explanation of the solar cycle modulation of the irradiance by observed \( dT \) is physically linked to the explanation by a varying bright network as indicated by e.g. the HeI index is not yet clear, but we note that the two interpretations are not necessarily inconsistent with each other. A more detailed study of the amplitude variation of p-modes may shed some light on this question.

5 Conclusions

The past decade has seen the introduction of radiometers with sufficient precision and sampling to measure a variety of small variations of the total solar irradiance. These variations are interesting to solar physicists in several ways. In addition, the presence of a distinct 11-year modulation of total irradiance suggests that longer-term variations may be significant for the Earth’s climate.

For most of the variations at least qualitative explanations are available; the details of the underlying physics, however, is not yet fully understood and more simultaneous studies of the detailed features on the solar disk together with continued solar constant observations are needed. A major result from the present data sets is that the solar cycle modulation of the irradiance is a true luminosity change of the Sun which can be modelled without assuming a simultaneous radius change of any significant magnitude.
Acknowledgments: As part of this paper and some Figures are from a review recently prepared for the "Sun in Time", a book to be published by the Arizona University Press, Tucson (Fröhlich et al. 1990b); I would like to take the opportunity to thank the co-authors Peter Foukal, John Hickey, Hugh Hudson and Richard Willson for their contributions, yet unpublished data and many helpful discussions. Thanks are extended to Jeff Kuhn for many helpful discussions and for providing material before publication. Furthermore, the continuous support of this work at PMOD/WRC by the Swiss National Science Foundation is also acknowledged.

References

Andersen, B. (1990): "Simulation of time series for granulation, super- and mesogranulation", private communication


