UNIVERSITY OF SOUTHAMPTON

FACULTY OF SCIENCE

School of Physics and Astronomy

Reconstruction of solar irradiance variations for use in studies of global climate change: Application of recent SoHO observations with historic data from the Greenwich observatory.

By

Simon Steven Foster

Thesis for the degree of Doctor of Philosophy

June 2004

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

SCHOOL OF PHYSICS AND ASTRONOMY

Doctor of Philosophy

RECONSTRUCTION OF SOLAR IRRADIANCE VARIATIONS, FOR USE IN STUDIES OF GLOBAL CLIMATE CHANGE: APPLICATION OF RECENT SoHO OBSERVATIONS WITH HISTORIC DATA FROM THE GREENWICH OBSERVATIONS.

by **Simon Steven Foster**

The Greenwich record of sunspot observations forms a unique, continuous data set spanning from May 1874 to December 1976, and as such is an extremely important resource for research into century-scale solar variability. The Mt. Wilson data provides a valuable extension of this dataset to the present day; however there are inter-calibration difficulties, which must be resolved. These issues are particularly important because there are several valuable modern datasets (for example magnetogram data and space-based observations of the total solar irradiance, TSI), which commence shortly after the cessation of the Greenwich data and thus can only be compared with Mt. Wilson observations. For studies of Earths climate systems we need to reconstruct solar variability, and in particular TSI, over the past 150 years and this requires the use of the Greenwich data, which must therefore be rigorously calibrated with respect to the Mt. Wilson data. In this thesis, I develop a composite of observations from a variety of datasets that is homogeneous in both sunspot area and sunspot position.

I use this continuous and homogeneous dataset to investigate known long-term variations in sunspot latitudes and area, and relate these to variations in the open magnetic flux of the Sun, as recently discovered from analysis of historical records of geomagnetic activity. The dataset is also readily applicable to the reconstruction of the darkening effect of sunspots on TSI, but reconstruction of the brightening effect of faculae is more challenging. I investigate the implications for facular brightening if the correlation of open flux with TSI were to apply on century as well as decadal timescales. This work has implications for paleoclimate studies which use cosmogenic isotopes as an indicator of solar variability.

Lastly, I employ recent SoHO observations of the contrast of flux tubes such as faculae in conjunction with the composite sunspot data series to model the variation in facular brightening over the past century, and using inferred variations in the solar magnetic field. This procedure allows the brightening by active region faculae around sunspots to be reconstructed with some certainty, but still requires assumptions to be made about the behaviour of network faculae and any other small flux tubes outside active regions. Three models of TSI are developed, based upon three assumptions. These models are compared to the standard models and satellite measurements of TSI, and the possible implications for the detection and attribution of the various factors, which have influenced global climate change over the past 150 years are discussed.

Contents:

List of Figures:

List of tables:

Acknowledgements

Firstly I would like to thank my mum, dad and sister for all their love and support (both financial and emotional) over the past 25 years, I feel this work is a culmination of all our efforts and hard work, and I know I could never have achieved all I have without you. I am sure at certain points, where I am today and what I have achieved would have seemed impossible, but thanks to you it has become reality! From an early age you got me interested in space and the solar system, I can still remember looking at the craters on the Moon through a telescope in the garden of Hobbs close, which is probably where all this started! I would also like to thank my Grandparents and extended family for all their help over the years, sadly not all of which here today to share this period of my life with me, but their influence, love and memory have shaped me into the person I am today. A big thank you must go out to my girlfriend Miss Emma Reynolds for all her love and support over the last couple of years. Thankfully she has been on hand to help me through the rough patches I have encountered, as well as share in the happy times, of which I hope there are many more. I would also like to thank my friends who I have met along the way, for all the happiness they have given me.

On the work side of things, I must first thank Prof. Mike Lockwood, for all his guidance, knowledge, patience, time and help that he has given me over the past 4 years. I can only say thank you for giving me the opportunity to actually undertake a PhD, something that I never thought possible. I would like to thank the STP group at the University of Southampton for all their help, and for putting up with me! I would like to thank the STP community as a whole, for all their help in answering my questions, and making the data used in this thesis freely available to me. I would also like to take this opportunity to thank Dr. R. Showan, for his help and encouragement he gave me whilst at school, sparking my interest in physics, and for setting me on this path.

Lastly, but most importantly, I would like to thank my nephew Kian, just for being himself! He has brought a great deal of joy into all our lives, and I can only hope we all continue to share in his life for many years to come. Thank you and good luck.

Some Words of Warning:

"As we sat there a blind man approached us. We learned afterwards that he had gone blind from gazing too long and too persistently at the Sun, trying to find out what it is in order to seize its light.

He strove a long time to accomplish this, constantly looking at the sun; but the only result was that his eyes were injured by its brightness, and he became blind.

Then he said to himself:

'The light of the sun is not liquid; for if it were a liquid it would be possible to pour from one vessel into another and it would be moved, like water, by the wind. Neither is it fire: for if it were fire, water would extinguish it. Neither is light a spirit, for it is seen by the eye; nor is it matter, for it cannot be moved. Therefore, as the light of the Sun is neither liquid, nor fire, nor spirit, nor matter, it-is nothing!'

So he argued, and, as a result of always looking at the Sun and always thinking about it, he lost both his sight and his reason. And when he went quite blind, he became fully convinced the Sun did not exist" *Leo Tolstoy*.

Chapter 1: Introduction.

1.1. Basic properties of the Sun and heliosphere.

The Sun is the source of the energy that powers our climate systems and allows life on Earth. At the same time, the Sun generates energetic particles, a solar wind and a magnetic field, which spreads throughout the heliosphere, the region of space dominated by the solar outputs. The heliosphere shields the Earth from galactic cosmic rays and energetic particles generated by supernova explosions. Table 1 lists some details of the Sun.

Table 1.1. The Characteristics of the Sun (SI units)

Solar radius, R_s (radius of the visible disk, the photosphere)	6.9599×10^8 m = $109.3R_E$ (an Earth radius, $1R_E = 6.37\times 10^6$ m)
Solar mass, m_s	1.989×10^{30} kg = $3.33 \times 10^5 m_E$ (an Earth mass, $m_E = 5.97 \times 10^{24}$ kg)
Surface area	6.087×10^{18} m^2
Volume	1.412×10^{27} m^3
Age	4.57×10^9 yr
Luminosity	3.846×10^{26} W
Power lost in solar wind	1.414×10^{22} W
Surface temperature	5770 K
Surface density	2.07×10^{-5} kg m^{-3}
Surface composition (by mass)	70% H, 28% He, 2% (C, N, O, ...)
Central temperature	1.56×10^7 K
Central density	1.50×10^5 kg m^{-3}
Central composition by mass	35% H, 63% He, 2% (C, N, O, ...)
Mean density	1.40×10^3 kg m^{-3}
Mean distance from Earth, D_{ES}	1.50×10^{11} m = 1AU = 215 R_s
Mean angle subtended by a solar diameter at Earth = $2\tan^{-1}(Rs/D_{ES})$	0.532°

Mean solid angle subtended by a solar disk at Earth = $\pi(Rs/D_{ES})^2$	6.7635×10^{-5} sr
Surface gravity	274 m s^{-2}
Escape velocity at surface	6.18×10^5 m s^{-1}
Equatorial rotation period (w.r.t. fixed stars)	25.5 days (frequency, f = 460 nHz)
Equatorial rotation period (w.r.t. Earth)	27 days
Polar rotation period (w.r.t. fixed stars)	34 days (frequency, f = 340 nHz)
Mass loss rate	10^9 kg s^{-1}
Inclination of equator w.r.t. ecliptic	7°

The visible solar surface is called the photosphere, which lies at an average heliocentric distance $r = R_s = 6.96 \times 10^8$ m. The regions below the photosphere are not directly observable and our knowledge of them comes from application of the helioseismology technique, from numerical models and, now that we understand more about their mass and oscillations, from neutrinos which can escape the interior without interacting. At the centre of the Sun lies the core ($0 < r < 0.25\ R_s$) where the high pressure and temperature cause the thermonuclear reactions, which power the Sun. The energy is then passed, mainly by the diffusion of gamma rays and X-rays, through the radiative zone ($0.25\ R_s < r < 0.7\ R_s$). Were it not to interact, a photon would cross the radiative zone in 2s; however it is scattered, absorbed and re-radiated so many times that this journey takes 10 million years. Above the transition region around $r = 0.7\ R_s$, the energy is brought to the surface by large-scale circulation across the convection zone, driven by buoyancy forces. The upflows and downflows are seen in the surface by the pattern of granules (of order 10^6 m across), which are organised into larger and deep-rooted circulation cells called supergranules (of order 10^7 m across, and also called the network) and giant cells (of order $0.3R_s$). The lower part of the solar atmosphere, the chromosphere, is 2.5×10^6m thick (so it covers $1R_s < r < 1.004R_s$). The temperature of the atmosphere increases dramatically at the transition region at the top of the chromosphere and is very high (of order 2×10^6K) throughout the main part of the solar atmosphere, the corona. As can be seen during eclipses, the corona has no clear outer edge; instead it evolves into the heliosphere, the region of space dominated by the solar wind outflow of ionised gas (plasma) and the weak magnetic field, also of solar origin, that is carried with it. One convenient threshold

that can be thought of a separating the corona from the heliosphere is $r = 2.5R_s$, beyond which the solar wind is no longer accelerated and the magnetic flux pulled out of the Sun is approximately constant.

The journey to the Earth takes solar photons 500s but the thermal charged particles of the solar wind take anything between about 2.5 and days. More energetic particles travel more rapidly, for example a 100MeV solar proton event (SPE) would take only 20min to reach Earth. The solar wind is slowed at a termination shock which theory predicts could be anywhere between $r < 10$AU and 120 AU, but is generally well beyond all the planets. Beyond this shock, the slowed solar wind continues to flow out to an outer boundary, the heliopause, the location of which could, in principle, vary in location between about $r = 50$AU and 150AU, depending on the pressure of the interstellar wind which meets the solar wind at this boundary.

The high temperatures in the solar interior mean that it is fully ionised. The primary ion is protons and in the core the high pressure of the overlying layers, and the high temperature overcome Coulomb electrostatic repulsion, and press pairs of protons together

$$^1H^+ + {}^1H^+ \rightarrow {}^2D^+ + \beta^+ + \nu_e \quad (+1.44\text{MeV})$$

where the superscripted number gives the atomic mass number of each reactant. The products are a deuterium nucleus, a positron and an electron neutrino and 1.44MeV of energy. This reaction proceeds only relatively slowly, but is followed by two further reactions:

$$^2D^+ + {}^1H^+ \rightarrow {}^3He^{2+} + \gamma \quad (+5.5\text{MeV})$$

where γ is a gamma-ray photon:

$$^3He^{2+} + {}^3He^{2+} \rightarrow {}^4He^{2+} + {}^1H^+ + {}^1H^+ \quad (+12.9\text{MeV})$$

Then nett effect of this chain is to convert 4 protons into a helium ion, with a mass loss of $\delta m = (4m_H - m_{He}) = 0.029$ amu. Thus the energy released is:

$$\delta E = \delta m\, c^2 = 27\text{MeV} = 4.3259 \times 10^{-12}\text{J}$$

where c is the velocity of light.

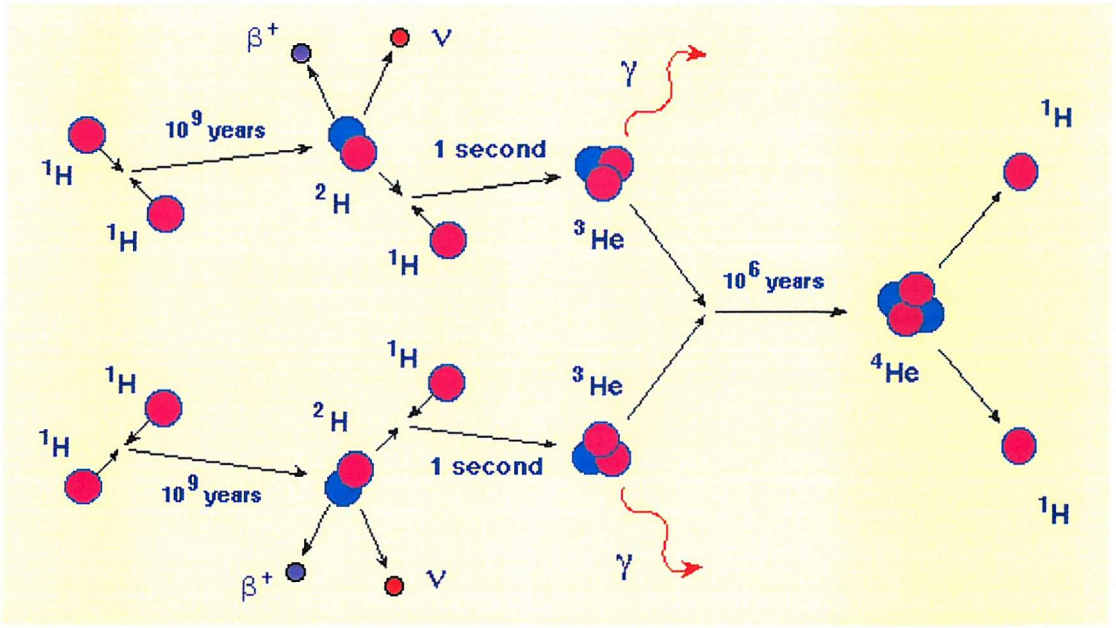


Figure 1.1. The proton-proton chain, the reaction chain liberating 27MeV in the Sun's core by converting 4 protons into a Helium ion.

To supply the present-day luminosity of the Sun (L = 3.846 × 10^{26} W, see Table 1.1) requires the reaction chain to be completed N = (L / δE) = 9×10^{37} times per second, for which protons in the core are lost at a rate of 3.6×10^{38} s^{-1} and helium ions are generated at 9×10^{37} s^{-1}. Thus the proton mass that is used up is 6×10^{11} kg s^{-1}. The above chain of nuclear reaction is called the proton-proton chain and is not the only one active. It is, however, the most important.

The Sun has long held a special wonderment in the human mind, with many ancient civilisations worshipping the Sun as a god, which isn't surprising since our existence depends it. Apart from nuclear energy (fission and fusion) all energy sources on Earth are the result (directly or indirectly) of the Sun. The possibility of nuclear fusion as a power source on Earth is in its infancy, but it is the fusion of hydrogen into helium inside the Sun which has provided the Sun's energy for billions of years. To estimate the lifetime of the Sun, studies are made of heavy radioactive elements found in meteorites, the Moon, and here on Earth, which although only make up a tiny fraction of the reactions inside the solar interior, are found in great enough quantities to allow us to estimate that the Sun has been luminous for approximately 5 × 10^9 yr [*Taylor*, 1997].

In comparison, given that the central composition of the Sun is 35% protons and the core presently contains of order 1.2×10^{28} kg of hydrogen, the rate of hydrogen burning by the proton-proton chain discussed above would yield a hydrogen burning phase of length 2×10^{16} sec = 6.4×10^{8} yr. By this time the Sun will have swollen to roughly twice its present radius before collapsing down to about 1.1 R_s for a helium burning phase lasting for a further 10^9 years [*Schröder et al.*, 2001]

The Sun is our nearest star and as such is the easiest star to study. Two separate scientific groups have evolved; solar physicists who wish to understand the structure and evolution of our star and stellar-solar astronomers who wish to place our star in context by comparison with many Sun like stars.

1.2. Historical observations.

The first recorded solar observations were made around or before 800 BC in China, and can be found in the 'book of changes'. These solar observations refer to large dark regions that are visible upon the solar surface, and must have been remarkably large to have been visible with the naked eye (it is speculated that these dark regions were visible through the smoke from forest fires). Astronomers at the Imperial Chinese court were under instruction to chart the movement of the astronomical bodies, so that the emperor could be made aware of changes in the heavens, and so these strange features on the Sun's surface would have been of extreme importance, and they were believed to be dragons emerging and diving into the solar surface. It is now obvious that these records were referring to large Sunspots visible on the solar surface, but at the time these now relatively common solar features would have been treated with a great deal of reverence and interest.

The first actual Sunspot drawing, was made in 1128, by John of Worcester, a monk, who observed the Sun with his naked eye under suitable viewing conditions, such as clouds, mist etc. It amazing to see and read the detail with which the Sunspot was observed, in his drawings both the umbra and penumbra are visible. The Sunspots must have been incredibly large to have been visible with such detail.

The real scientific study of Sunspots began after the telescope had been brought into astronomy in 1609, and there is some controversy as to whom the discovery of Sunspots should be accredited to. Four astronomers can lay claim to the discovery of Sunspots and they were Johann Goldsmid (a.k.a. Fabricus) in Holland, Thomas Harriot in England, Galileo Galilei in Italy and Christopher Scheiner in Germany. Both Harriot and Galileo were the first to record Sunspot observations, and the oldest properly recorded telescopic observation of a Sunspot was made by Harriot on the 8th of December 1610, and the entries can be seen in his notebooks. However these observations were not published and he did not continue his observations in any continuous fashion. Fabricus was the first to publish his results in 1611, and he even interpreted the motion of the Sunspots in terms of the rotation of the Sun. It was Galileo who first proposed that Sunspots were a feature of the solar surface, as opposed to planets close to the solar surface. This proposition that the Sun's surface could be blemished by dark features was against everything that the views of the Roman Catholic Church of the day, which held that the Sun was a perfect sphere, and so Galileo's views on Sunspots contributed to his appearance before the Roman Inquisition in 1633.

Even though Sunspots had entered the modern, western literature in 1611, they were not studied systematically until the late part of the 17th century at the Paris observatory. Even then investigations into Sunspots were slow to gain any momentum due to the lack of Sunspots on the solar surface at this time. This period was investigated by Maunder in the 19th century and he noted that this period was highly unusual, and this time of low Sunspot activity has come to be known as the Maunder minimum. The period of the Maunder minimum was noted by Eddy to coincide with a period of extremely low temperatures in the northern hemisphere, and was called the little ice age. The first serious discussion of such a link was by Herschel in 1801, and there have been many statistical studies, often of no statistical significance and beset by selection effects, but viable mechanisms have only emerged over the last decade.

The Royal Greenwich observatory commenced Sunspot observations on the 9th of May 1874, and this was the first time that a consistent record of Sunspot position, number and area had been undertaken. Thanks to the records at the Paris and Greenwich observatories, in the 19th century Schwabe showed that the number of

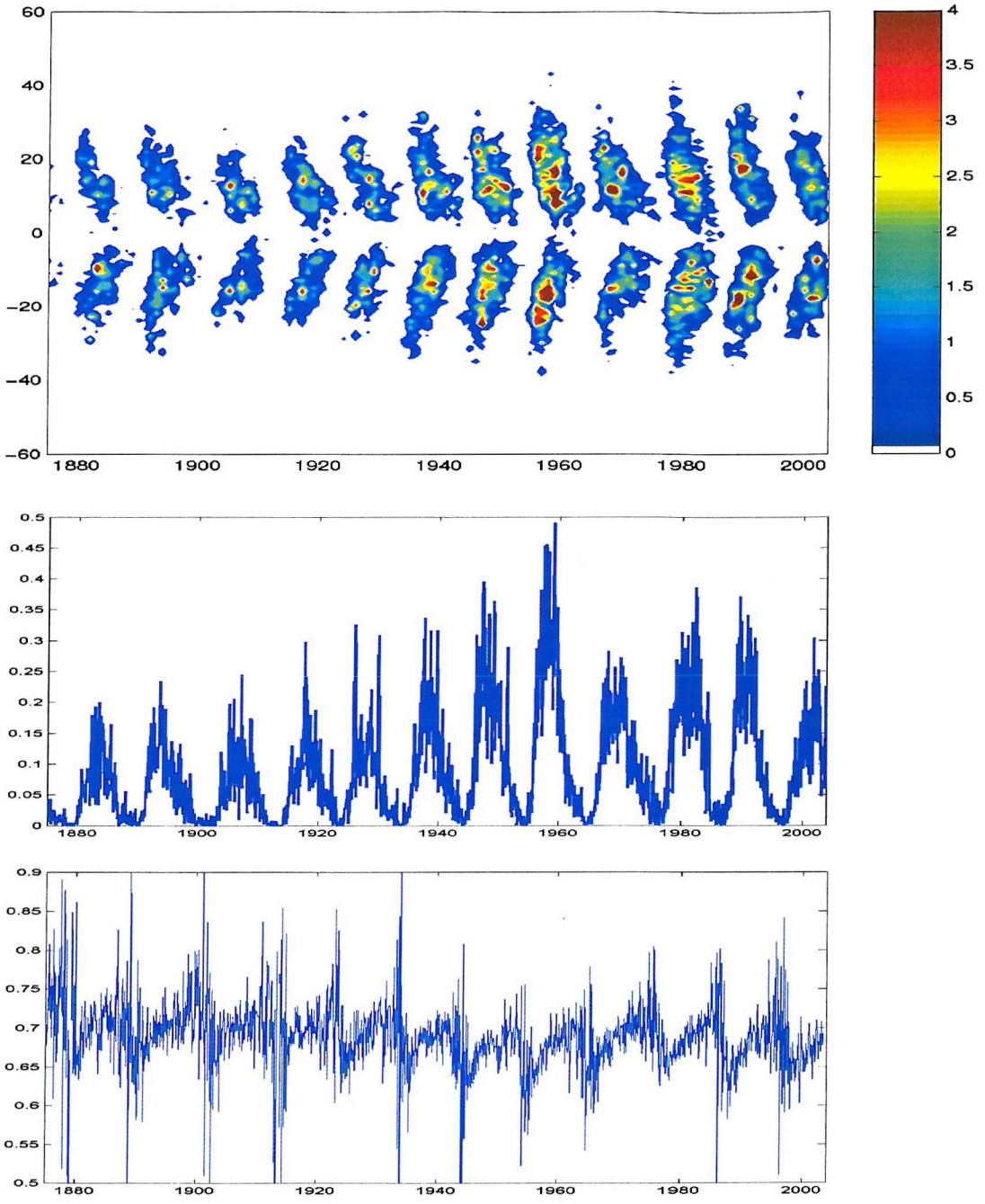


Figure 1.2. Top panel: Contour plot of annual percentage solar surface covered by spots in 1° latitude bins. Middle panel: Monthly mean percentage of the total solar surface covered with Sunspots. Bottom panel: Monthly mean position of Sunspots (μ) with respect to the centre of the solar disk.

Sunspots was periodic, with a period of approximately 11-years between the successive times of minimum spot number. Carrington then showed that the position of Sunspots during the solar cycle migrate in latitude, with spots being at high solar latitude at the start of the cycle, and being at low solar latitudes by the end of the cycle. This behaviour is displayed in the famous butterfly diagram, which can be seen in figure 1.2.

The solar observations continued at the royal Greenwich observatory until 1981, where the Mt. Wilson observatory took over solar observations. Chapter 5 of this thesis deals with the intercalibration of these two datasets. The combined dataset has proved invaluable in solar studies of the long-term variations and changes in the Sun. With these records solar researchers have been able to see the variations of solar latitude, area and number of spots, and have thus have been able to gain a better understanding of the Sun. These records are at the core of Sun-climate links presented in this thesis.

Hale made a very important discovery in the field of solar physics in 1908, using the Zeeman splitting of spectral lines. In the absence of a magnetic field several quantum mechanical states of an atom can have the same energy, even though they have different quantum numbers. A magnetic field can destroy this symmetry of the system, by causing a preferred direction and there is splitting in the energy levels, if a spectral line is now observed which occurs from the transition involving the states concerned, it is split into several components. *Hale* used this principle to observe Sunspots and found that Sunspots are regions of strong magnetic field and usually occur in bipolar pairs, and hence this is why Sunspot groups are sometimes referred to as bipolar magnetic regions (BMR's). It also became evident from *Hale*s work that magnetic polarity of the leading spot in the bipolar pair (in terms of the solar rotation) is opposite in the two hemispheres and changes from one solar cycle to the next, so the Sun has a 22-year magnetic cycle. Discoveries followed from this showing that the whole solar magnetic field flips every eleven years. It has been shown that magnetic flux of the trailing spot polarity migrates pole ward over the course of the solar cycle away from the solar equator. As this flux accumulates in the polar 'coronal holes' it causes a reversal of the polarity of the polar field. This behaviour can be seen in figure 1.3. Large magnetic fields are concentrated into two latitude bands, which define the butterfly diagram, and this is the region in which solar magnetic filed emerges through the solar surface, I will refer to these bands as the 'active regions' throughout this thesis. The dispersed flux, which is left over from an active region, is

called ephemeral flux. The magnetic flux that has been dragged out into the heliosphere by the solar wind is called open and this forms the coronal holes. Surface flux is moved round supergranulation cells as they form and fade. This causes flux to accumulate in the dark lanes of these cells, giving a structure called the network.

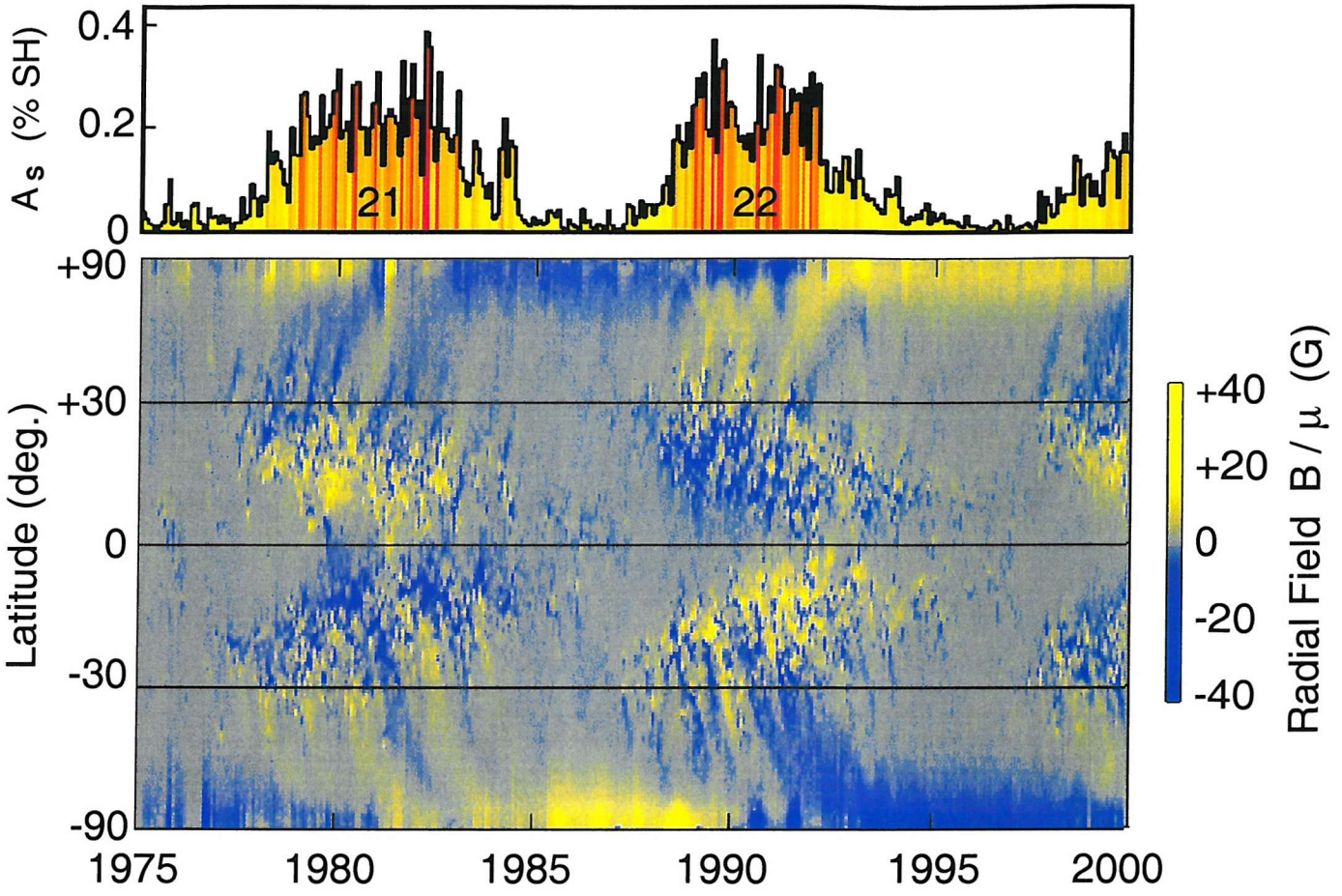


Figure 1.3. The association of Sunspots and magnetic field. (Top) The total area covered by Sunspots (A_S - given in % of the visible solar hemisphere) for 1975 to 2000, covering solar cycles 21-22. (Bottom) Longitudinal averages of the radial field (B/μ) at a function of latitude (positive northward) and time where B is the line-of-sight field observed by magnetographs (positive outwards) and $\mu = \cos\theta$ where θ is the heliocentric angle. [Courtesy of *NASA/Hathaway* 2003, http://science.nasa.gov].

Granules are small (about 1000 km across), ubiquitous cellular features that cover the entire Sun except for those areas covered by Sunspots. These features are the tops of small convection cells where hot fluid rises up from the interior in the bright areas, spreads out across the surface, cools and then sinks inward along the dark lanes that are about 400° cooler. Individual granules last for only 18 minutes on average. The granulation pattern is continually evolving as old granules are pushed

aside by newly emerging ones. The circulation flow speeds within the granules are typically 1 km s^{-1} but can reach supersonic speeds exceeding 7 km s^{-1}. The granulation of the quiet photosphere can be seen outside the Sunspot in figure 1.4.

Sunspots are the visible manifestation of concentrated magnetic fields at the solar surface. These magnetic fields come from deep within the solar interior and emerge through the solar surface. Where the magnetic field emerging is strong enough it can inhibit the transport of heat flux from deeper in the solar interior to the surface, this then causes the temperature within the effected solar region to drop and so become darker than the surrounding region of the photosphere. Sunspots are made up of two regions, the darker central region called the umbra, and the somewhat lighter surrounding region called the penumbra. Spot sizes vary greatly, but typical umbral and penumbral diameters are 20×10^6 m and 40×10^6 m. On average Sunspots are around 2000K cooler than the normal photosphere, and the magnetic induction in the Sunspot umbra is about 0.3T, falling to approximately 0.15T in the penumbra. The magnetic field is approximately vertical in the umbra and it becomes increasingly more horizontal through the penumbra. Recently it has been observed that surrounding Sunspots a faint bright ring can be observed, and this could be due to heat flux that is diverted around the flux tube and appears at the surface, this additional heat flux would then cause a rise in the surface temperature surrounding the spot [*Rast et al.*, 1999]. Sunspots typically last for several days, although very large ones can survive for several weeks.

As mentioned previously Sunspots have been observed at various observatories around in the world, such as India, U.S.A, the former USSR, but by far the longest continuous observations have been made at the Greenwich observatory in the U.K. When skies were cloudy or it was otherwise impossible for solar observations to be made at Greenwich, data from other observatories, such as India and Mauritius were utilised. Measurements of position (heliographic longitude, latitude, and distance from centre of solar disk), area (both umbral and whole spot), number and group types were all conducted at these observatories. These observations spanned the period 1874 – 1976, when the series was discontinued. Equivalent observations have since continued at the Mt Wilson observatory in the USA.

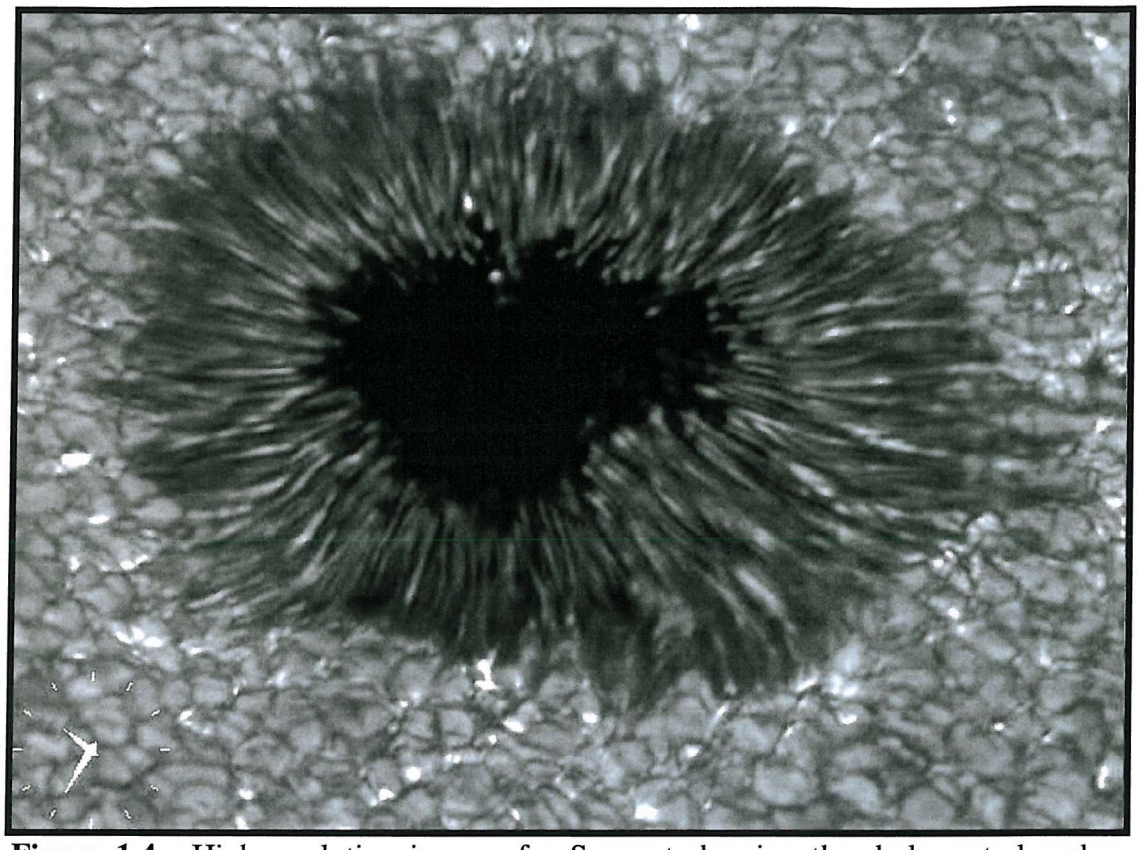


Figure 1.4. High-resolution image of a Sunspot showing the dark central umbra, filamentary penumbra and the granulation of the photosphere surrounding the spot. [*Royal Swedish Academy of Sciences,* www.solarviews.com]

Sunspots are not the only solar features associated with magnetic flux threading the solar surface. Faculae are small bright regions on the solar surface caused by regions of enhanced magnetic flux, which are smaller in diameter than those that cause Sunspots. Their name originates from the Latin meaning 'little torches'. As in Sunspots, the strong magnetic fields inhibit the transport of heat flux to the surface, but whereas in Sunspots this causes a darkening and lowering of the temperature in the feature, faculae are found to be brighter and hotter than the surrounding quiet Sun, and this difference is associated with the smaller size of the features. The material inside a facular magnetic flux tube is heated from the walls surrounding the flux tube. This causes the temperature within the faculum to be the same as the surrounding solar surface. There are different theories as to why faculae are bright. The most frequently used idea is that the plasma pressure within the flux tube will be lower than the surrounding regions because of the enhanced magnetic

pressure. This means that the optical depth within faculae is lower and so it is possible to see deeper within the solar interior where it is hotter and denser. In particular it is possible to see the bright wall, especially near the limb where the faculae are viewed at a large angle to the vertical. Faculae are not restricted to the active region and can occur somewhat randomly outside the region, thus faculae are split into two separate groups, those that occur within the active regions (these are called active region faculae), and those that occur all over the Sun in the lanes of the supergranulation cells (these are called network faculae). Network faculae are usually smaller than active region faculae and although they follow the solar cycle variation in number, like active region faculae, they do not show a similar butterfly distribution in their latitudes.

Measurements of faculae were initially made at the Greenwich observatory for an intermittent period between 1886 –1917. These faculae measurements were white light measurements, and as such only observations close to the limb of the solar disk were possible. In early observations, telescopes could not distinguish faculae from the background quiet Sun in the disc centre, but as the faculae rotated out towards the solar limb, their contrast increased with respect to the surrounding solar surface. This was partly due to the hot wall effect discussed above but also because of the limb darkening effect on the quiet Sun. Limb darkening arises because near the limb the light has to pass through a longer slant path through the solar atmosphere. This causes the unit optical depth to be raised and thus we see light from higher in the photosphere, where it is cooler and less dense and so emits less intensity. This is the opposite in faculae, where the lower density causes the unit optical depth to be lowered, and thus we see deeper into the photosphere. Thus faculae are now distinguishable from the surrounding solar surface and so their position and size can be measured. In recent times, observations of faculae, such as those made at the San Fernando observatory, who have employed ground-based photometric images made using a red filter (λ = 672.3 nm), and images obtained in the Ca II K line (λ = 393.4nm), were studied and any pixels that were brighter than the surrounding photosphere by a value of 4.8% were classified as faculae and their areas were calculated [*Chapman et al.* 1997].

The Ca II K along with the Mg II doublet is used as a proxy for a number of brightening features on the solar surface, such as faculae and plages. The Ca II K absorption feature is centred at a wavelength of 393.4 nm, and is the result of the overlap of two Ca II absorption lines originating in the solar photosphere. Two Ca II K emission lines at the base of this composite feature originate above this in the solar chromosphere. A similar effect is evident in the Mg II absorption line variation, with the Mg II line being found in the UV spectrum. The variations of this feature with respect to the variable core and the more stable limb are used to define variations in the photosphere and chromosphere. Magnetic plages (from the French meaning beach) are bright structures that occur in the chromospheric layer of the Sun's atmosphere. They are prominent in the Ca II line, and the total magnetic flux of plage regions is comparable to that of Sunspots, and their behaviour is similar to that of Sunspots [*Foukal*, 1996]. Plage radiation is one of the main sources of the EUV radiation from the Sun, and plages appear bright in respect to the quiet Sun and have been used as a proxy for solar faculae. Plages are more extensive than faculae because the small emerging magnetic flux tubes increase in cross sectional area with height above the solar surface.

1.3. Surface and open magnetic fields

The magnetic fields in active regions and the quiet Sun can be seen with the aid of a magnetogram. The colour coding on each magnetogram image depends upon the observatory, as can be seen in figure 1.3 yellow areas show positive magnetic fields (pointing towards the observer), blue show negative magnetic fields (pointing away from the observer), and grey regions show areas of no detectable magnetic field. In other magnetograms the corresponding colour coding is white, black and grey. Note the importance of pixel size in a magnetogram. For a large pixel equal and opposite strong fields can cancel to yield a low (near zero) value. Magnetograms made it clear that Sunspots appear in magnetic pairs, one of negative polarity and one of positive polarity, and these magnetic poles are joined by magnetic loops that rise up into the solar atmosphere and arch like bridges between the bipolar groups. In this thesis I make use of the Michelson-Doppler Imager (MDI) magnetogram on the SoHO satellite, which has allowed more detailed studies of the solar magnetic fields to be undertaken.

The rate at which magnetic flux is emerging through the active regions is called the flux emergence rate (E), and this varies over the solar cycle, and the total emergence rate has been modelled by *Solanki et al.* [2000] in order to fit the variation in open solar flux derived from terrestrial observations by *Lockwood et al.* [1999a]. Active region flux emerges quickly at the solar surface and it is destroyed from the solar surface due to interactions between the opposite polarities in bipolar groups. These interaction occur at the borders between the opposite polarity groups and this region is called the polarity inversion line (PIL). [*Solanki et al.,* 2000, *Mackay et al.,* 2002]. As mentioned Sunspots appear in bi-polar pairs, and the leading spot usually occurs at a lower latitude, i.e. towards the solar equator. This was seen by *Hale* [1919] who noted a tilt in the east-west direction, and this tilt is measured in terms of the tilt angle of the axis of the bipolar magnetic group relative to the azimuthal direction. Since there is a difference in the rotational rate depending upon solar latitude lower latitude spots rotate at a faster rate, and so the leading spot will move away from the trailing spot. This will cause the PIL to be stretched out causing the destruction of the magnetic flux to take place over a longer time frame. Thus the emergence rate is variable depending upon the latitude, tilt and area of the Bipolar magnetic group.

Not all magnetic flux emerging form the solar surface is destroyed through reconnection with magnetic flux of the opposite polarity. Some magnetic flux continues to rise up through the solar atmosphere and becomes embedded in the solar wind outflow, which drags it out into the heliosphere. This 'open flux' forms the heliospheric field, which in the vicinity of the Earth is also called the interplanetary magnetic field (IMF). One definition of 'open flux' is when the solar magnetic field becomes approximately radial, and this is thought to occur at a heliocentric distance of approximately 2.5 R_S (where R_S is a solar radius). This is called the coronal source surface, and the magnetic flux threading this surface is called the 'coronal source flux' or the 'total open flux', F_S. On annual timescales, the interplanetary magnetic field (IMF, the heliospheric field in the ecliptic plane), obeys the Parker spiral orientation with a garden hose angle γ, which is approximately constant at a given radius. The IMF's radial component near the Earth can be measured directly or estimated from the geomagnetic aa index. In addition the Ulysses spacecraft has shown that the radial component of the heliospheric field is independent of latitude if we arrange over

longitudinal structure [*Lockwood et al.*, 1999b]. Thus if the radial component of the IMF near Earth has magnitude $|B_r|$, then the total signed flux threading a heliocentric sphere of radius R = 1AU is $F_s = 4\pi R^2|B_r|/2$. the factor 2 arises because half of the radial flux is outward and half is inward. *Lockwood* [2002b] has shown that the flux threading a sphere of radius 2.5 R_s is very similar to that described by this equation. From this equation *Lockwood et al.* [1999a] derived a method for calculating the coronal source flux from the geomagnetic aa index.

Solar activity causes geomagnetic activity through energy extracted from the solar wind flow by the Earth's magnetosphere. This has been monitored in a sustained and homogeneous method since 1868 using antipodal observing stations in England and Australia. The level of activity was quantified by *Mayaud* [1972], who developed the aa index, using the range of the fluctuations of the horizontal component of the field observed in 3-hour intervals at these stations. The aa index was designed to be very well correlated with other planetary indices of geomagnetic activity, such as the planetary *Ap* index [*Mayaud,* 1972]. The long-term change in the aa index is found to be highly correlated with related terrestrial phenomena such as the occurrence of low latitude aurora [*Pulkkinen et al.* 2000]. The aa data sequence shows a systematic long-term rise in activity since 1900, and follows a similar rise to the variation in number of parameters, Sunspot number and emergence rate and the decline in cosmogenic isotopes deposited in ocean sediments, ice sheets and tree trunks. Intriguingly the rise also mirrors that in global average temperature on Earth [*Cliver, 2002*], and this has been interpreted as showing global warming is totally solar in origin. In fact 'detection attribution' techniques reveal that although there is a detectable solar influence on Earths surface temperature, most of recent the recent variation are due to volcanoes are due to volcanoes and man-made greenhouse gasses [*Stott et al.*, 2000]

1.4. Cosmic rays

The Earth is shielded from galactic cosmic rays by the heliospheric magnetic field, and variations in this magnetic field are found to be strongly anti-correlated with cosmic ray fluxes incident upon the Earth's atmosphere [*Lockwood*, 2001]. The field is dragged out of the Sun by the continuous, but variable solar wind and fluctuates with the magnetic cycle of the Sun. This results in the fluxes of those cosmic rays that

do penetrate the shield and reach the Earth, showing a clear solar cycle variation with higher cosmic ray fluxes when Sunspot numbers are lower. There are three classes of particles called cosmic rays:

1. Galactic Cosmic Rays (GCRs). These are accelerated at the shock fronts of explosive galactic events such as supernovae explosions. The flux incident on our heliosphere is expected to vary on very long timescales as our Sun passes through the spiral arms of our galaxy. There is some evidence from meteorites for this long timescale variation. The bulk of these particles have energies between about one and a few tens of GeV. The flux of GCRs has been continuously and systematically monitored now for over 50 years by a network of ground observatories, which measure the neutrons or muons that GCRs produce in the atmosphere. Earth's magnetic field adds to the shielding caused by the heliosphere and limits the ranges of particle that can be seen at any one site. The key parameter is the rigidity of the GCRs, which is a measure of the tendency of the particle to keep moving in a straight line. Rigidity is measured in GV, but because the particles move at close to the speed of light, their energy is close to the rigidity value expressed in GeV. The geomagnetic field places a cut-off threshold on the particles that can be seen at any one site. Close to the geomagnetic equator this cut-off is close to 15GeV, whereas at mid latitudes particles of a few GeV and above can be seen. At the polar latitudes the geomagnetic cut-off falls below a 1-2 GV cut-off set by the atmosphere.

2. Solar Cosmic Rays. These are generated at the shock fronts of explosive events on the Sun, for example the leading edge of Coronal Mass Ejections (CMEs). Because they are somewhat lower energies (up to several hundred MeV) and come from the Sun, these are now generally referred to as *solar proton events* (SPEs).

Svensmark and Friis-Christiensen [1997] and *Svensmark* [1998] have reported a solar cycle variation in the global fraction of terrestrial cloud cover seen from space. They propose that this is due to cloud cover being directly influenced by the galactic cosmic rays incident upon Earth. Concerns were raised due to the compilation of the

cloud data because it was compiled from a diverse set of sources, and there is also no clear dependence upon the type of clouds that is controlled by cosmic rays, which would be expected. Subsequent work by *Marsh and Svensmark* [2000], found that there is strong correlation between cosmic ray fluxes and low altitude clouds, and so is mainly associated with stratus and stratocumulus cloud types in maritime regions.

With the Earth's atmosphere being constantly bombarded with cosmic rays, radioactive isotopes are produced, with ^{14}C and ^{10}Be being amongst the most important. Both ^{14}C and ^{10}Be are produced as spillation products of GCRs colliding with O, N and Ar gas in the atmosphere. About 2/3 is produced in the stratosphere and 1/3 in the underlying troposphere. Thereafter the deposition of the two isotopes is very different. ^{10}Be attaches to aerosols and precipitates into ice sheets where its abundance can be measured in an ice core. Dating is done by counting the seasonal layering of clean and dirty ice and by modelling the flow of the ice sheet. Some key dates are given by dust from known volcanic eruptions (e.g. Krakatoa). ^{10}Be is also deposited and measured in ocean sediments. On the other hand, ^{14}C becomes CO and CO_2 and is exchanged with two major reservoirs, the biomass and the oceans. Understanding the production rate of ^{14}C from records of its abundance found in, for example, tree trunks, requires complex modelling of the exchange of ^{14}C between the atmosphere and the biomass and ocean reservoirs. As mentioned previously the rate of incident cosmic rays is anti-correlated with the heliospheric magnetic field and as such the production of ^{14}C and ^{10}Be will vary with the heliospheric magnetic field. Variations in the production of the ^{14}C and ^{10}Be isotopes, can be used to measure the past variation in the solar magnetic fields. This has been investigated by *Lockwood* [2001] who found the ^{10}Be isotope to be an excellent proxy for coronal source flux variations and has constructed the variation in coronal source flux back to 1400. The data *Lockwood* used was obtained from ice core drillings in Greenland, and show that by the end of the Maunder minimum the solar magnetic field was extremely weak, lending more weight to the speculation that the Sun was inactive during this period.

1.5. Total solar irradiance.

The Sun's influence on the Earth's climate is not limited to any possible effect of GCRs. The Sun has a more direct input into the Earth's climatic system, in the

form of the Sun's energy output called the total solar irradiance. The total solar irradiance (TSI) describes the radiant energy emitted by the Sun over all wavelengths that falls each second on 1 square metre normal to the Sun-Earth line outside the Earth's atmosphere. TSI is measured in Watts per square meter. It must be remembered that the total solar irradiance applies to a particular orbital position and distance from the Sun, due to the fact that the Earth's orbit is not quite circular. In addition, the solar output is directional and so values seen at Earth are not precisely the same as would be seen at a heliocentric distance of 1 AU out of the ecliptic plane and over the solar poles. The Sun can be modelled as a blackbody with an effective temperature of approximately 5770K. the power flux emitted by unit area of a blackbody is $F = \sigma_{SB}T^4$ where σ_{SB} is the Stefan-Boltzman constant ($5.6696 \times 10^{-8}Wm^{-2}K^{-4}$). Thus for the effective surface temperature of 5770K, F is 6.284×10^7 Wm^{-2}. If the temperature is constant over the entire surface of the Sun, the total solar luminosity would be $L_s = 4\pi R_s^2 F$ (where R_s is the solar photospheric radius = 6.96×10^8m). This yields $L_s=3.825 \times 10^{26}$W. If the power is radiated isotropically the TSI would be $I_{TS}=L_s/(4\pi R_1^2)$, where R_1 = 1AU = 1.5×10^{11}m. this yields a TSI of 1353 Wm^{-2}. It was impossible for the solar irradiance to be measured accurately from the Earth's surface because the Sun was viewed through Earth's atmosphere. The atmosphere, although transparent to most frequencies of solar radiation, is highly absorbent in the ultraviolet and infrared frequencies, and so an accurate measurement could never be taken from the Earth's surface. With the advent of the space age, accurate measurements of the Sun's irradiance could be undertaken. In November 1978 the Nimbus 7 satellite was launched, and one of its many tasks was to investigate the solar irradiance, and subsequent satellites have been launched with instruments to continue this investigation (ACRIM I and II, ERBE, SoHO). The Nimbus 7 and following satellites used radiometers to measure the solar irradiance, and although these radiometers varied from satellite to satellite, the basic design remained the same. A radiometer consists of a number of pyroheliometers, which each consists of 2 cavities, and temperature differences between the two are used to determine the total solar irradiance. One of these cavities is maintained at a constant reference temperature, whilst the other is heated 0.5K higher than the reference cavity and is exposed to the Sun. When the shutter covering the second cavity is opened, Sunlight enters, creating an even greater difference in cavity temperatures. The power supplied to the second cavity decreases automatically to maintain the 0.5K

temperature difference between the two cavities. This decrease in the amount of electricity is proportional to the solar irradiance entering the cavity.

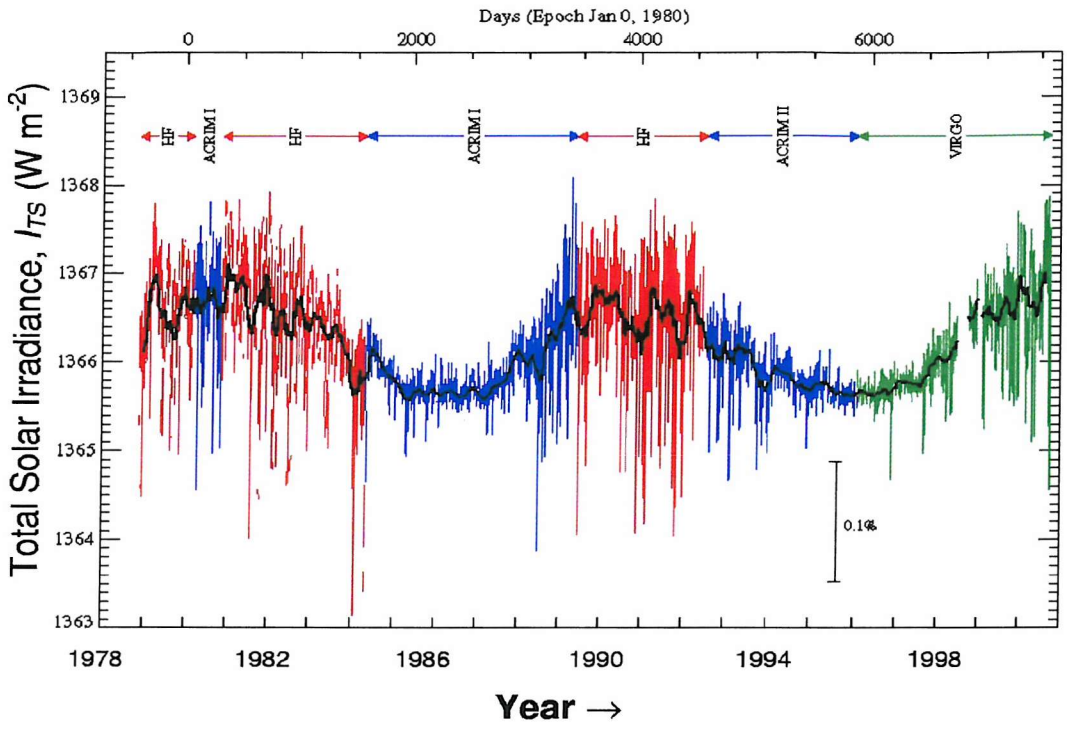


Figure 1.5. Composite of several datasets from different spacecraft showing the total solar irradiance variation since 1979. Daily values are shown in red, blue and green (for HF, ACRIM 1-2, and VIRGO, respectively), monthly means in black (from *Fröhlich and Lean*, 1998).

The findings of the Nimbus 7 satellite proved to be startling. The solar sensor found that the total solar irradiance varied over the solar cycle by approximately 0.1%, with a similar waveform to Sunspot number, with the largest amount of TSI being found at solar maximum, and smallest amounts at solar minimum. For many years the reverse effect had been expected. At solar maximum Sunspot number is found to be greatest and thus the amount of Sunspot darkening would be largest, and so an overall decrease in irradiance would be expected if this were the only effect. The difference is mainly due or wholly the result of faculae. As was discussed earlier, faculae are bright features on the solar surface and their numbers are found to vary over the solar in a similar fashion to the Sunspot number. Thus the effects of one prove to counter the effects of the other. At solar minimum, numbers of both features are small and their effect on solar irradiance is small, but as magnetic activity increases through the solar cycle both Sunspots and faculae increase in size and

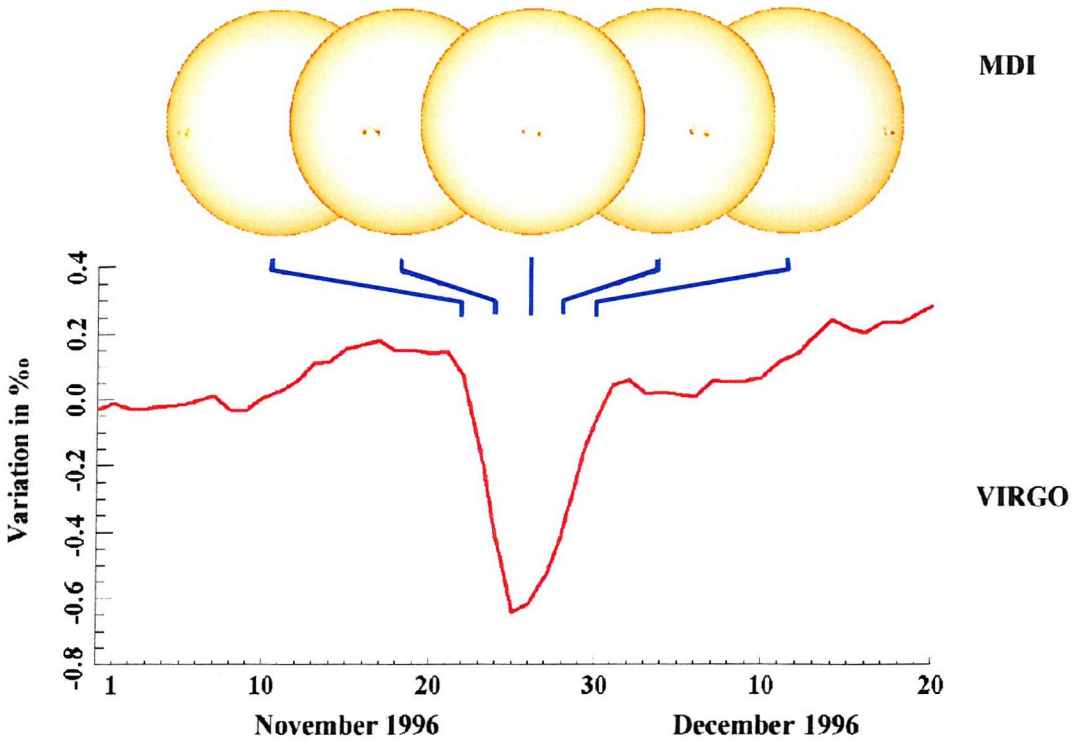


Figure 1.6. The effect on TSI of a Sunspot group rotating across the visible disc.

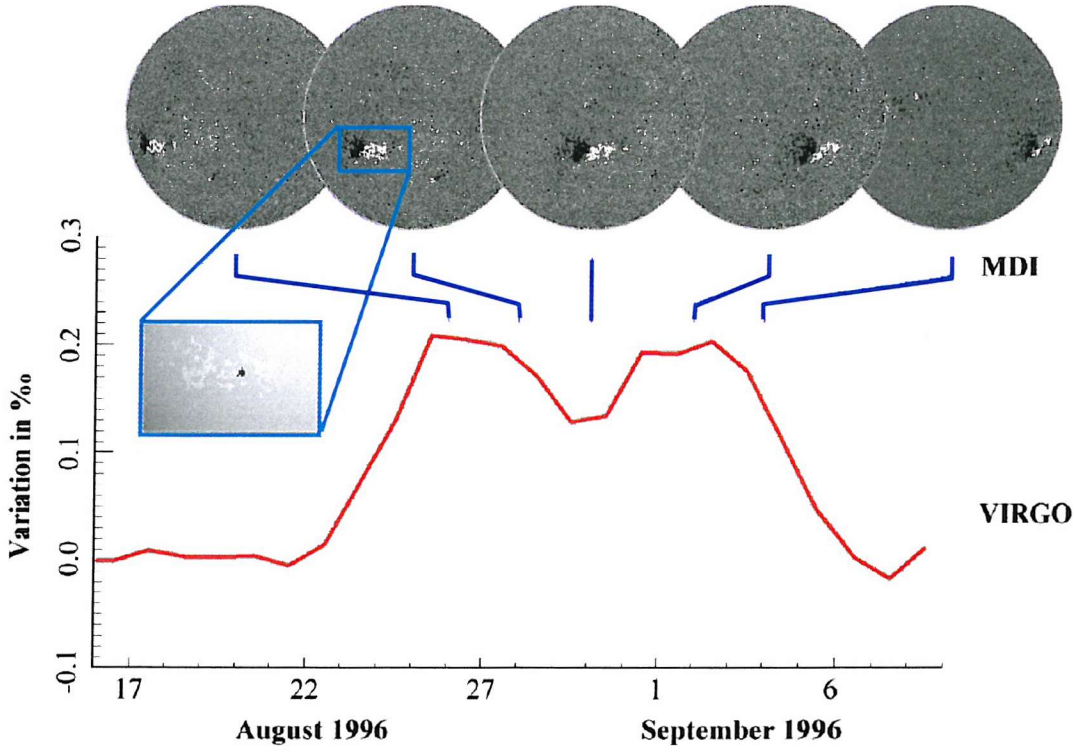


Figure 1.7. The effect on TSI of a faculum, which can be seen in magnetograms from SoHO as it rotates across the disc. The inset shows the continuum emission. Black and white show inward and outward field and grey is field that is below the detectable threshold.

number, and it has been found that the area covered of the Sun by faculae is roughly ten times the area covered by Sunspots. Increased solar magnetic activity actually leads to increases in both brightening and darkening by the relevant features, and it has been found that the brightness associated with faculae outweighs the darkening effect of Sunspots, thus leading to an overall increase in irradiance at Sunspot maximum, relative to Sunspot minimum.

To study the effect of Sunspots upon the solar irradiance, the photometric Sunspot index (PSI) has been developed, so that the loss in irradiance can be quantified [*Frohlich et al.*, 1994], and it was found that PSI varied over the solar cycle by around 0.1% of TSI. It has been found that variations in PSI are dominated by changes in to the fraction of the solar disk covered in Sunspots. This makes it relatively easy to reconstruct the variation in PSI. In the opposite respect, the brightening due to faculae (f_b) is relatively unknown. Faculae have proved harder to resolve on the solar surface and so it has proved harder to quantify their effect on total solar irradiance, this is due to faculae being smaller in size and thus monitoring their contribution with respect to the quiet Sun has proved difficult. Since the role of Sunspots in irradiance variations is relatively well understood, defining the present and past role of faculae on solar irradiance variations has proved crucial to understanding the behaviour of the total solar irradiance in the past. It is believed that active region facular brightening varies with active region facular area, and thus recent facular brightening can be fairly well reconstructed using the facular area measurements, for example *Chapman et al* [1997, 2001]. Unfortunately facular area measurements are not available to make any long-term reconstruction of facular brightening, and so assumptions need to be made to model past facular behaviour. The effect of network faculae on the total solar irradiance have proved even more difficult to quantify, due to the relatively unknown behaviour of network faculae, and although network faculae define only a fifth of total facular effects on the solar surface, their contribution to long-term irradiance variations could be of great importance.

As noted previously, when looking at the limb of the solar disk, as opposed to the centre, we see light that has taken a slanting path through this layer and this gives "limb darkening" as we only see the upper, cooler and dimmer regions. This has two

opposite results, in Sunspots, limb effects will cause Sunspots to appear slightly darker, whereas in faculae limb effects will cause faculae to appear brighter. This is best explained when we look at variations in contrast at the disk centre and limb. For Sunspots the contrast remains relatively constant at around 0.3 with respect to the quiet Sun, whereas in faculae the contrast varies from 1.01 to 1.1 at disk centre and limb, with respect to the quiet Sun. Thus faculae will have the greatest effect upon TSI near the limb, and this effect can be seen in figure 1.7. On the other hand, Sunspots have a peak darkening effect when at the centre of the solar disc (see figure 1.6). Thus the position at which solar features appear and spend their existence could play an important role in solar irradiance variations.

By careful intercalibration of the various TSI data from the various irradiance monitoring satellites, the variation of TSI since the 1978 is extremely well understood (see figure 1.5). It should be noted that two separate composites of TSI have been developed (called the PMOD (Physics and Meteorological Observatory Davos) and ACRIM composites respectively), by *Fröhlich* [2000] and *Willson and Mordinov* [2002], and differences arise due to different interpretations of the satellite data sets. This is investigated in further detail in chapter 2. Both reconstructions show that minima TSI has remained fairly constant over the past 2 solar cycles, but to understand what impact TSI variations have had upon global climate change, a robust long-term (century-scale) TSI reconstruction is needed. This is because the large thermal capacity of the Earth's oceans largely damps out decadal variations in the terrestrial response to TSI changes: however, it will not damp out century-scale variations. Various models of the long-term TSI variations have been made, using a variety of methods. *Lean et al.* [1995] and *Lean* [2000a] used a combination of Sunspot number and 11-year running means of Sunspot number to develop a model of TSI back to 1700, *Solanki and Fligge* [2000] developed a TSI model using a complex combination of Sunspot number, area, and Ca II K line plage images, and *Hoyt and Schatten* [1993] used the solar cycle length. In all these reconstructions, the amplitude of the century scale drift in TSI was set by assuming that non-cyclic Sun-like stars were in Maunder-minimum state and so the TSI during the Maunder minimum could be set from statistical studies of these stars. In this respect, the reconstruction by *Lockwood and Stamper* [1999] is radically different as it was an attempt to avoid use of a scaling parameter based upon stellar analogues. *Lockwood and Stamper* reported

a strong relationship between annual values of F_s and total solar irradiance (TSI), which then allowed annual TSI variations to be reconstructed back to 1868. The cause of this connection between TSI and F_s is not fully understood, but it can be speculated, that since variations in the solar magnetic field drive variations in Sunspot and faculae, then variations in F_s, could be linked to the causes of irradiance variations.

The important distinction between total solar irradiance and solar luminosity, L must be stressed. The total solar irradiance depends upon the position from which solar observations, but solar luminosity is the amount of energy per unit time emerging from the entire solar surface, thus luminosity may not vary in the same fashion as TSI. Variations in TSI caused by Sunspots and faculae seen in the ecliptic plane, such as at Earth, may have a completely different effect upon TSI incident upon a body in another plain of orbit, say above the solar poles. So as TSI incident upon the Earth increases towards solar maximum, it may be decreasing in the plain of orbit above the poles, in which case it is impossible that solar luminosity may show no change over the solar cycle, or even in the past century. Since we have never observed the Sun from over its poles we have no observations of L, but we can obtain a value from theory of 3.845×10^{26}W.

All the TSI reconstructions discussed above show a long-term increase in TSI over the past 100 years, and if these reconstructions are to be believed, this TSI increase could have profound effects upon the global temperature. From 1860 to 1990, the mean global annual temperature increased by 0.55°C, and at the same time concentrations of CO_2 gas in the Earth's atmosphere increased from 280 to 353 ppmv, leading to the conclusion that the warmer global temperatures show the climate systems response to anthropogenic influences. However it has been found that the climate also shows signs of interannual and interdecadal variability, which can be linked to solar variations, suggesting that the cause of warming is more complex than a direct just a simple link to increasing greenhouse gases. The global climatic system is a highly complex structure, with natural influences such as volcanic activity and solar variability not being fully understood. The Earth's oceans are capable of levelling out short-term variations (around 10 years), which is why solar cycle variations in solar irradiance have not been found to significantly influence global

temperature, and to this end, long-term variations in the quiet Sun irradiance have proved of great interest. It is also difficult to try and quantify anthropogenic effects due to the fact that as industrial activity has increased over the past 100 years, so has solar activity, and so the solar influence needs to be accurately quantified so that the anthropogenic effect, if any, can be isolated and properly quantified. As has been mentioned previously *Svensmark and Friis-Christiensen* [1997] and *Svensmark* [1998], proposed a link between cosmic rays and global cloud cover, and if correct, this is another effect upon the global climatic system that needs to be fully understood and investigated to properly quantify variations in the global temperature. The role clouds play in the global climate system is important, clouds increase the global albedo, leading to a higher percentage of solar irradiance being reflected into space, thus leading to a decrease in surface temperature. However, the effect is not straightforward as clouds also trap heat and can lead to warming.

As would be expected, the question of the magnitude of the impact of solar variability on the global climate is a highly complex and contentious one, and to form an answer, solar processes need to be understood, and long-term behavioural trends in a number of solar parameters such as irradiance and magnetic activity need to be understood and modelled successfully. To this end, in this thesis I have studied this area of research and have developed various models of the TSI, PSI, facular brightening, emergence rate and coronal source flux, with the aim that they can be used to gain better understanding of our space environment, Sun, and global climate system. My aim is to avoid both unexplained correlations and uncertain stellar analogues, but instead to use recent understanding of the effects of solar magnetic fields, in particular as obtained using the SoHO spacecraft.

Further introductory material will be given in chapter 4, where a review of the *Spruit's* theory of sunspot blocking shall be presented.

1.6. Goals of this thesis.

In this thesis it is my aim to develop accurate reconstructions of the total solar irradiance based upon the physics of the variations, rather statistical correlations, where the physics linking the proxy and irradiance are not well understood. To this

end I break TSI down into its components parts, faculae (network and active region), sunspots and quiets sun, and each parameter is investigated separately. Models of each of these parameters will be presented and there impact upon TSI shall be assessed. It is also my aim to avoid the use of stellar analogues, which have been utilised in previous irradiance reconstructions, as questions are now being raised as to their effectiveness at defining long-term irradiance variations in the Sun.

I also aim to investigate the link between open solar flux (and therefore cosmic rays and cosmogenic isotopes) and TSI. Paleoclimate studies have utilised cosmogenic isotopes to reconstruct the long-term variation in TSI, due to similarities in the variations in the two parameters over the last 2 solar cycles, yet a physical mechanism between the two is not yet fully understood. It is my aim to verify that cosmogenic isotopes can indeed be used to define the long-term variation in TSI, by comparison with the independent models I develop.

Chapter 2: Overview of data sets used.

2. 1 Chapter overview.

In this chapter, I shall explain the various data sets that will be used in this thesis. Many of the data sets used are composed of composites from various sources, and the intercalibrations need to be discussed carefully so results derived from their use can be fully understood and of optimum scientific reliability.

Because we are interested in long data series (greater than 5 solar cycles), it is very rare that a data set is compiled from a single source, such as a single observatory or satellite. Instead the data series must usually be compiled from a variety of sources, and even when a data set is collected from one source, there are usually problems encountered with continuity and consistency inside the data sets. For example, as mentioned in chapter 1, the composite data set of Total Solar Irradiance (TSI), is compiled from a variety of different instruments onboard satellites, each of which gives slightly different values of TSI and each subject to different drifts. The results from these different instruments are intercalibrated, with a single continuous data set being constructed from the various sources. Even when a composite is constructed, it does not rule out the factors such as individual interpretation of the data, which can lead to variations in the data sets. In the case of the TSI composites this individual interpretation has lead to different composites being constructed. These variations in data set composition will be discussed and the explanations given for the use of data sets in this thesis.

This dependence upon individual interpretation of results can also be seen within the Greenwich series of sunspot observations. Continuous observations of sunspot groups have been made at Greenwich since 1874, but concluded in 1976. Although the Greenwich data are as homogeneous as can be possible, being compiled using set procedures with allowance for improvements in the instrument sensitivity and resolution (see section 2.3.2), the extension to recent years is not straightforward. In particular for 5 years after the Greenwich data ceased some poor quality data has

been introduced into the data series. This is particularly problematic because there is no valid overlap to allow the intercalibration with Mt. Wilson observations. In this thesis I shall make use of a third dataset, recorded at the USSR network of solar observatories, which gives overlap with both the Greenwich data and the stable, high quality Mt. Wilson data available from 1982 onwards. This data has not been exploited previously in this context. The high confidence that can be placed on the composite derived using all 3 datasets is particularly important when one considers that TSI data are available from 1978 onwards and so overlap exclusively with the Mt. Wilson data, whereas extrapolation to century timescales would rely on the Greenwich data.

In this thesis, I aim to interpret long-term datasets using data and understanding from more recent datasets. In section 2.3, I discuss the longer-term data series, but first I examine the relevant recent data.

2. 2 Recent data sets.

2. 2. 1 Total Solar Irradiance (TSI).

Composite data sets of total solar irradiance have been constructed from a variety of satellite-based instruments, with differences in the interpretation and intercalibration of the measurements leading to a number of different composites being developed. The instruments used are:

- The Hickey-Frieden radiometer (HF) onboard the Earth Radiation Budget (ERB) experiment, from November 16th 1978 and lasted until January 24th 1993.

- ACRIM I (Active Cavity Radiometer for Irradiance Monitoring) monitored solar irradiance onboard the Solar Maximum Mission Satellite (SMM) from February 14th 1980 until June 1st 1989.

- The Earth Radiation Budget Satellite (ERBS) began solar irradiance monitoring from October 25th, 1984 and continued monitoring until December 1996. There was a 4 months gap in 1993 due to power problems onboard the satellite.

- ACRIM II onboard the Upper Atmosphere Research Satellite (UARS), from October 1991 until December 2000.

- SOVA on the European Retrievable Carrier (EURECA) from August 11th, 1992 until May 14th 1993.

- VIRGO (Variable of solar Irradiance and Gravity Oscillations) on the SoHO (Solar Heliospheric Observatory) satellite since January 18th 1996 until the present day. Small intervals of data were lost in 1998 due to loss of control of the whole SoHO craft.

- ACRIM III experiment, launched in December 1999, and began monitoring TSI in April 2000.

Each instrument varied in its design, and this of course would lead to variations in the observations due to technical capabilities of the instruments. Some instruments are/were self calibrating (generally using 2 radiometers, one routinely and one used sparsely) others are not. Descriptions of the instruments used can be found at www.nasa.gov and in *Fröhlich* [2003].

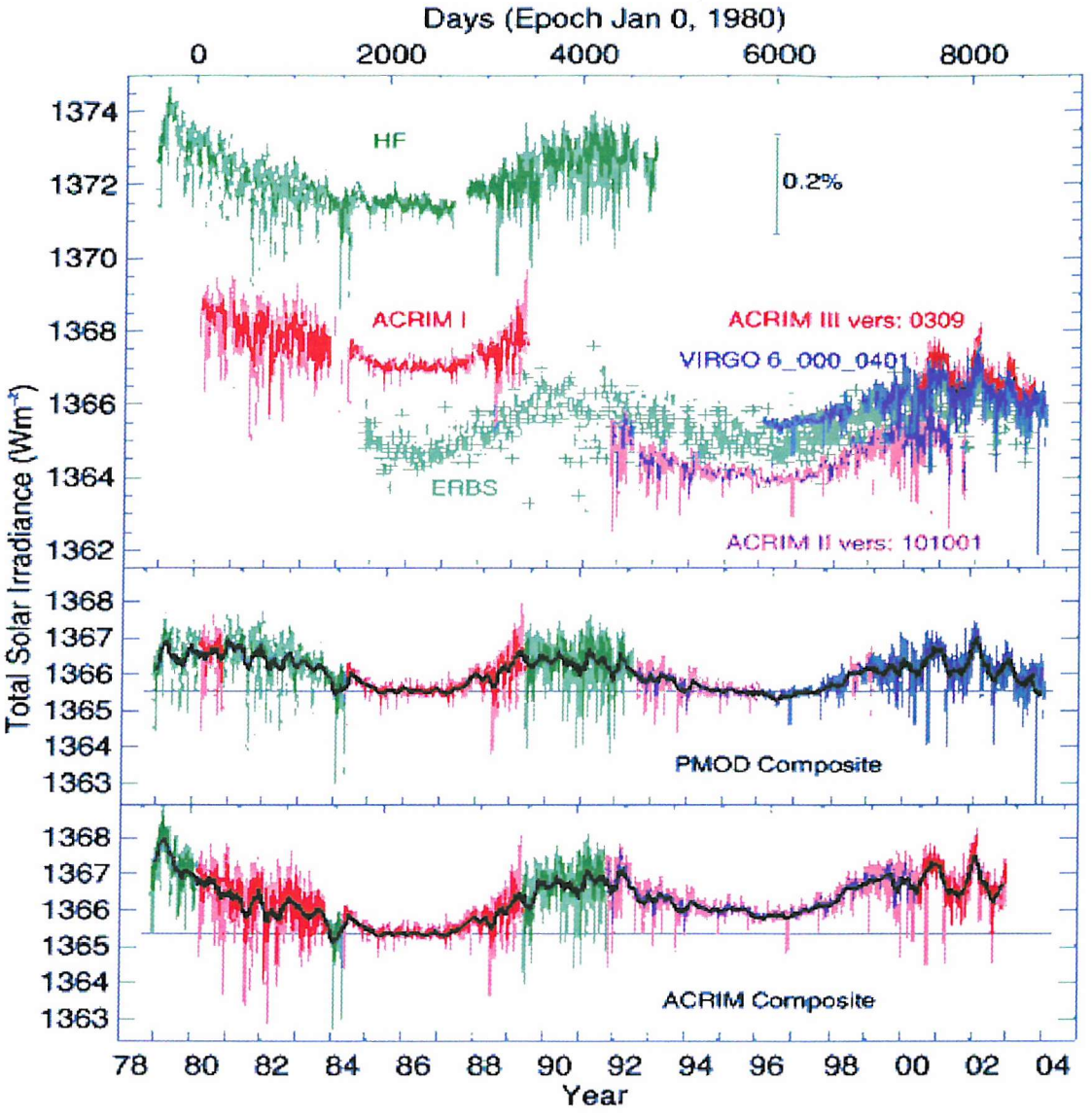


Figure 2.1: Top plot shows the various radiometers used to construct composites. Bottom plots show the main composites constructed from the top panels data. [http://www.pmodwrc.ch/solar_const/solar_const.html]

The data series that I have used in my work is the PMOD composite data series, constructed by Claus Fröhlich at the World Radiation Centre, Davos. The PMOD data set is a composite of the HF, ACRIM I, ACRIM II and VIRGO instruments and can be seen in figure 2.1. Corrections have been made for: the early measurements from the HF radiometer to account for degradation, the early degradation of ACRIM I, comparison of ACRIM II to ACRIM I, using ERBE, HF, assessments made for the influence of the many operational interruptions with the ACRIM II data series.. The full description of the construction of the composite can be found in *Fröhlich and Lean* [1998] and *Fröhlich* [2000]. It should be noted that the PMOD composite varies to the composite constructed by Dr. R. Willson at the centre

for climate systems research [*Willson,* 1997, *Willson and Mordinov,* 2002]. A comparison of the two models can be seen in figure 2.1, and there are a number of obvious differences between the two models. *Willson* suggests that the 1996/97 minimum is about 0.5 Wm^{-2} higher than the minimum in 1986/87, whereas the PMOD composite indicates a decrease of 0.13 Wm^{-2}. There is also a noticeable difference at the start of the two data series, it can be seen that the *Willson* composite continues to increase towards the start of the data series, whereas the PMOD composite peaks in 1981, and then decreases towards the start of the data series. This has been noted previously and the reasons for this are investigated in *Fröhlich and Lean* [1998a].

The key difference between the two composites is a glitch in the Nimbus-7 HF data caused by a satellite pointing change in 1990. Such a glitch had occurred previously and the PMOD composite is corrected for using the ERBS data. On the other hand, the ACRIM composite does not allow for this glitch and so requires a rapid and uncharacteristic degradation of the ERBS instrument at the same time [*Fröhlich,* 2003].

I here use the PMOD composite in the preference to the ACRIM for three reasons. 1: The NIMBUS-7 glitch is expected but extra-rapid degradation of ERBS is not. (2): Using PSI and facular proxies can match the PMOD composite but not the ACRIM composite [*Fröhlich,* 2003]. (3): Reconstructing irradiance variations from Kitt Peak magnetograms and a 5-component irradiance model matches the PMOD composite but not the ACRIM composite [*Wenzler, private communication,* 2004].

2. 2. 2 Photometric Sunspot Index (PSI).

Sunspots are dark regions of the solar photosphere caused by the concentration of magnetic fields causing blocking of the up flowing heat flux from lower down in the solar interior. The PSI data series I have used in this thesis has been constructed by Claus Fröhlich at the Davos observatory. The calculations of PSI utilises the records of sunspot area observations from around the globe, intercalibrated and published in the Solar-Geophysical catalogue to quantify the darkening effect caused

by sunspots upon the TSI. The fundamental equation in the calculation of the PSI data series is:

$$PSI = \sum_{S} \mu \frac{3\mu + 2}{2} a_s \alpha \tag{2.1}$$

where μ defines the position of the feature on the solar surface, a_s is the area of the sunspot in parts per millionths of a solar hemisphere and α is the contrast of the spot relative to the surrounding photosphere [*Fröhlich et al.,* 1994]. Contrast will be explained in further detail in chapter 4 of this thesis. The contrast is relative to an assumed "Eddington" limb darkening function. The sum is carried out over all spots present on the visible disc. Equation (2.1) takes into account the area dependence of the contrast and calculates the 'true' daily values for each observation using the differential rotation of the spots. Utilising equation (1), daily values of PSI in Wm^{-2} were calculated for the period December 1981 – December 1999 (see figure 2.2).

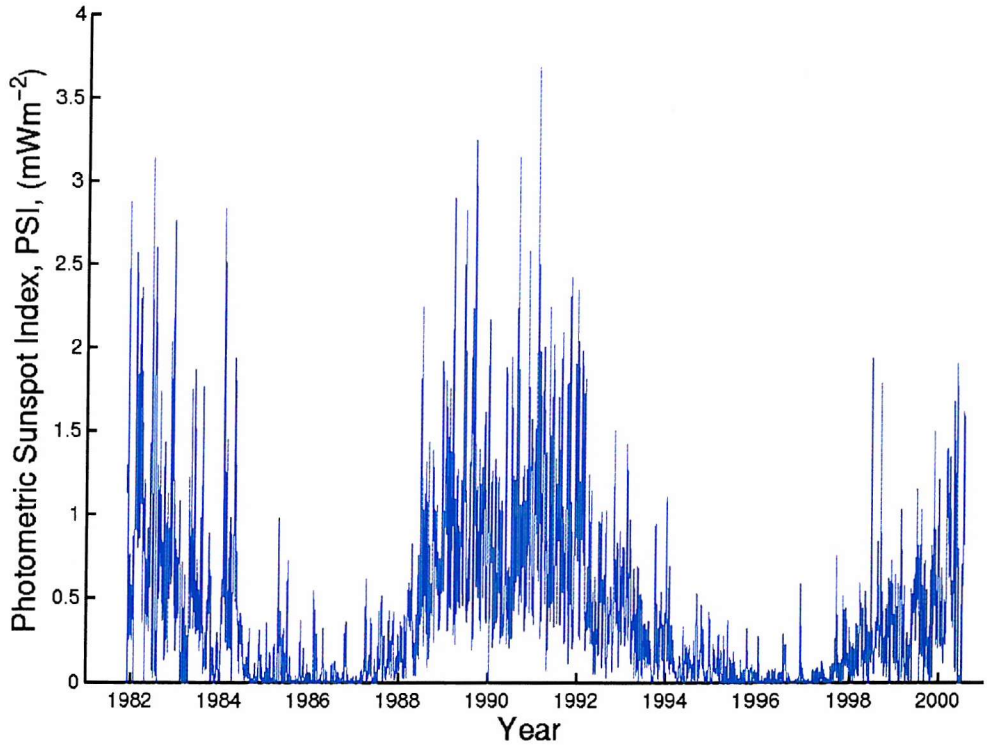


Figure 2.2: Daily values of the photometric sunspot index.

2. 2. 3. Facular Area.

Daily digital images of both faculae and sunspots have been made since March 2nd 1981 using the Cartesian Full Disk Telescope No.1 (CFDT1) at the San Fernando observatory, with a red filter (λ = 672.3 nm) and starting in May 1988, images were also produced in the Ca II K emission line (λ = 393.4 nm). The images produced, which include the sun and surrounding sky, contain 512 × 512 square pixels, each 5" × 5" and allow for objective measurements of facular areas.

Faculae are determined from the Ca II K line images, taken through the λ = 393.4 nm with a 0.9 nm bandpass. Facular pixels are identified from calibrated and flattened K-line images, as those which show a contrast that is equal or greater than a contrast criterion whose lower boundary is 4.8% at the centre of the disk [*Chapman et al.,* 1997]. The identified faculae pixels are then summed into daily values, which define the amount of the solar surface that is covered in faculae. Continuous daily observations have been made from May 1988 to the present day, with the results presented in millionths of a solar hemisphere (msh). This data was kindly provided by Prof Chapman at the San Fernando observatory [*Chapman, private correspondence,* 2003]. The data set of daily facular area derived from this method can be seen in chapter 6.

2. 2. 4. Cosmic rays.

The neutron monitor was developed by John Simpson at the University of Chicago in 1950. It was realised that by placing instruments at different latitude the Earth's magnetic field could be used as a crude spectrometer to allow measurements of the cosmic ray spectrum down to low energies. However differences in instrument sensitivity and the required corrections for atmospheric pressure mean that high resolution spectra of GCRs are not possible.

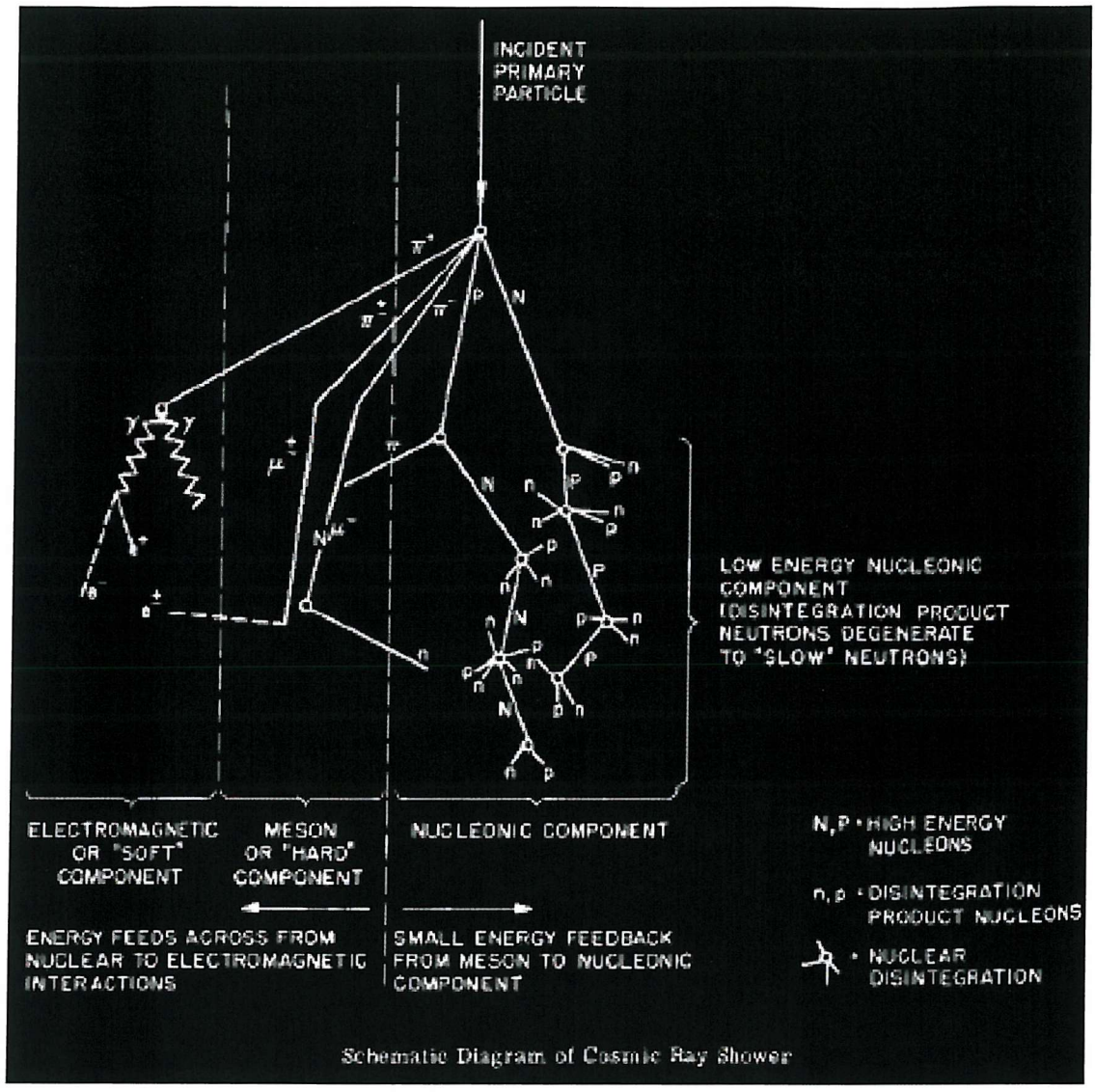


Figure 2.3: Diagram illustrating interactions of cosmic rays in the atmosphere (www.ulysses.sr.unh.edu).

In the early 1950's such network of high-altitude neutron monitor stations was established, covering a wide range of geomagnetic latitudes. The longest data series consists of the records taken at the Climax station in Colorado, USA (1951 – present day), and detects cosmic rays with energies greater than 3 GeV. The Huancayo station in Peru (1952 – 1992) measured cosmic rays with energies greater than 13 GeV. When political unrest made measurements in Peru impossible the observations were continued at the Hawaii station (1991 – present day). This station provided a close match to the geomagnetic latitude and altitude of the Huanncayo station, so that they both measured cosmic rays in the same energy band, i.e. greater than 13 GeV. The two monitors were operated simultaneously for twenty-one months, and so a sufficient amount of data was obtained to achieve a good intercalibration of the two

monitors and stations. Data from all cosmic ray sites are available on the ftp server of the World Data Centre for cosmic rays in Japan.

2. 2. 5. MDI: continuum emission intensities and magnetograms.

The Michelson Doppler Imager (MDI) on the SoHO (Solar and Heliospheric Observatory) is a CCD imager which images the Sun with 1024 pixels across a solar diameter and through a series of increasingly narrow spectral filters (Scherrer et al., 1995). The final elements are a pair of tenable Michelson interferometers, which allow MDI to record with a FWHM (full-width, half-maximum) bandwidth of 9.4 nm. Five wavelengths of the solar continuum emission near the Ni I solar absorption line at 676.8nm are recorded. The continuum intensity is computed using all five filtergrams obtained, the averaging reducing the observation errors. The MDI magnetograms are constructed by measuring the Doppler shift in right- and left-handed circularly polarised light, the difference between the two giving the Zeeman splitting which is exactly proportional to the magnetic flux density. The value obtained is the line-of-sight field and is an average over the pixel and noise levels are 20G, r.m.s. Intensity values have shot noise of order 0.3% and the flat-field intercalibration of pixels is good to within a few tenths of a percent or less. In routine operation ("FD mode") continuum intensity images and magnetograms have matching pixels of angular resolution 4″×4″ but higher resolution images ("HR mode") with 1.3″×1.3″ are available when the instrument is allocated a higher telemetry rate. These resolutions compare with 2.3″×2.3″ for the ground-based Kitt Peak Observatory magnetograms.

2. 3. Long-term data sets.

2. 3. 1. Coronal source flux.

As discussed in chapter 1, the coronal source surface is where the solar magnetic field becomes approximately radial and lies at a heliocentric distance of

about 2.5 solar radii. The total magnetic flux leaving the sun, and thereby entering the heliosphere by threading this surface, is called the coronal source flux, F_s. *Lockwood et al.* [1999a] have devised a procedure for calculating the coronal source flux F_s from the geomagnetic aa index. *Mayaud* [1972] devised the geomagnetic aa index to measure geomagnetic activity from a data series that extends in a continuously form back to 1868. The aa index is created from measurement stations in southern England and in Australia. Two things are striking from looking at the data, the first is the solar cycle variation that is evident, which is similar to other solar variables such as sunspot number, total solar irradiance etc. the second is the clear long-term increase in aa index over the past century. *Cliverd et al.* [1998] have conducted a study of the causes of the long-term increase, and found that it was due to a variation in interplanetary space, and thus the sun.

We can understand how the Sun can influence this drift if we consider how the aa index is derived. The range of horizontal perturbations to the geomagnetic field in every three-hour interval is scaled and converted into an aa index for each hemisphere (aa_N and aa_S) using the station-dependent scaling procedure devised by *Mayaud* [1972]. *Lockwood* [2003] looked at twenty year running means of the aa index from the Northern and Southern hemispheres, and found that the long-term drift is present in both hemispheres, but the Northern hemisphere is found to be 16% larger than the Southern Hemisphere. The same factor was found for the detrended solar cycle variations. *Lockwood* [2003] concluded that the underlying cause behind the solar cycle variation and the long-term drift must be the same in both cases, i.e. solar influence.

Figure 2.4 shows the annual values of *F*s derived by Lockwood *et al.* [1999a] from the aa index. It should be stressed that this procedure is not based upon a simple correlation. The calculation is based upon a combination of the theory of solar wind magnetosphere energy coupling, Parker spiral theory and an understanding of the role of fast solar wind streams in generating recurrent geomagnetic activity. This method is applied to twelve-month intervals to eliminate annual effects (such as the annual variation in the Earth-Sun distance, seasonal effects and the Earths dipole tilt). Twelve-month intervals can be advanced one month at a time to produce monthly data, but as a result of this, only every one in twelve estimates is truly independent.

This calculated open solar flux has been found to correlate extremely well with the annual means of open solar flux from the near-Earth IMF measurements [*Lockwood,* 2003].

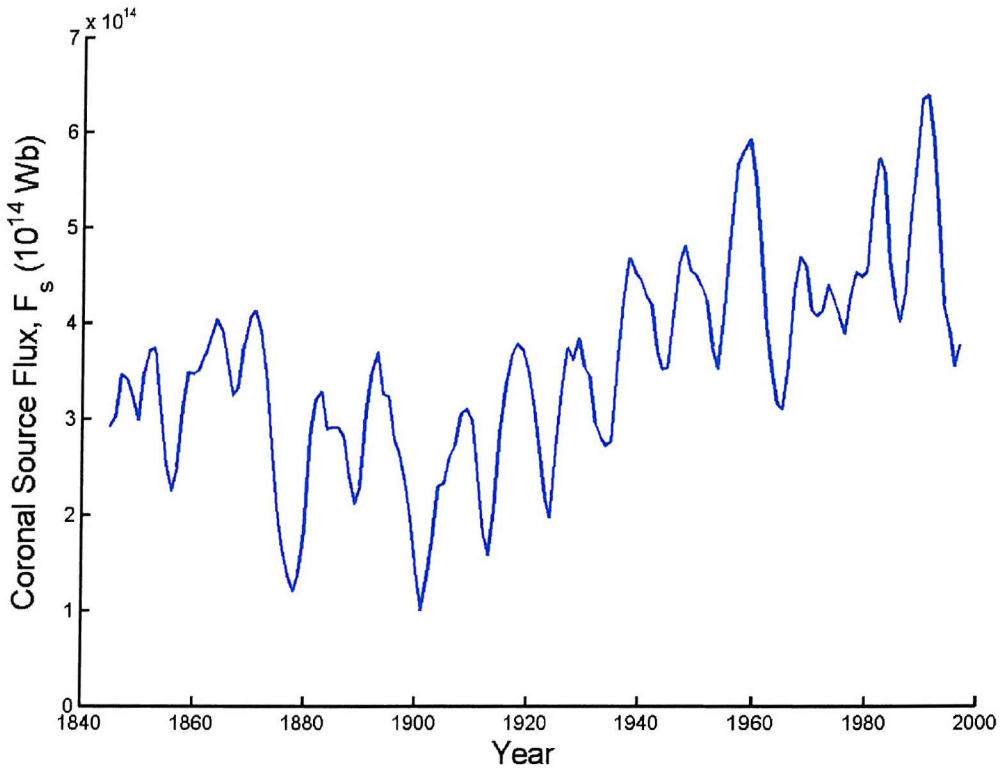


Figure 2.4: Yearly plots of the coronal source flux, F_s [data courtesy of *Lockwood*].

2. 3. 2. Greenwich, Mt. Wilson and USSR data.

Continuous solar observations commenced at the Royal Greenwich Observatory (RGO) on May 9th 1874, using a Dallmeyer telescope, with only a four-inch aperture and a focal length of five feet. The image of the Sun in the principle focus of the telescope was about six-tenths of an inch in diameter, but a magnifying lens is used, with the photograph actually obtained being approximately eight inches. Even with the great enlargement, the light of the Sun is so intense, that the exposure of the photographic plates can only occur for a fraction of a second. Various telescopes superseded this original telescope, but the general method used to obtain the photographs remained the same

Two of these photographs were taken every fine day at Greenwich, but due to the cloudy climate in the U.K, occasionally no good opportunity for photographing the sun occurred. In the winter, weeks could pass without the chance to obtain a photograph. When this occurred, photographs were taken with similar instruments in India and Mauritius (and sometimes in Hungary), and these photographs are sent to Greenwich when they are required to fill the gaps in the Greenwich data series. Therefore the Greenwich observatory provides a near continuous daily record of features on the solar surface.

Once a photograph of the Sun had been taken, it needs to be analysed, with the following parameters of each spot being calculated: firstly, its distance from the centre of the image of the sun; secondly, the angle between it and the north point; thirdly, the size of the spot; and fourthly, the size of the umbra of the spot. Observers measured the area of the spot by placing a thin piece of glass, on which a number of cross-lines have been ruled one-hundredth of an inch apart, on top the photograph. The cross-lines make up a number of small squares, each ten-thousandth of a square inch in size. When the photograph and plate were in contact, the number of squares covered by a given spot is counted, using a magnifying glass. The size of just one of these squares, in actual fact covers more than a million square miles on the solar surface. White light faculae were also measured in a similar fashion to sunspots outlined above.

After the solar photographs had been scaled, the features need to be reduced, and the positions of the features (either faculae or sunspots) expressed in longitude and latitude. This is relatively easily done, as the position of the sun's equator and poles are relatively well known.

The Greenwich data is presented with area, latitude, longitude, Greenwich sunspot number, Greenwich group type, distance from the centre of the solar disk and polar angle all given for each individual sunspot and faculae group. The Greenwich observations concluded on 31st of December 1975, and for the next five years observations from the Debrecken observatory in Hungary are used as a substitute for the Greenwich observations. These observations are made in a similar fashion to the original Greenwich data, although umbral and penumbral areas were no longer

measured separately. These data were simply appended to the end of the Greenwich data, to continue the data set to the 31st of December 1981. Recently concerns have been raised about the validity of some of the assumptions used to allow the continuation of the Greenwich data set using the Debrecken data series [*Hathaway, private communication*, 2004]. An early attempt to properly intercalibrate the Greenwich - Debrecken - Mt. Wilson appears to have been lost, and other attempts have been based on limited overlap data. Use of newly discovered data is investigated in Chapter 5.

Sunspot observations from the Mt. Wilson observatory in the U.S.A have been made from the 1st of January 1962 to the present day. The daily sunspot drawings are made using a one hundred and fifty foot solar tower at Mt. Wilson. The sunspot records are made using drawings in pencil on white paper. To make a drawing, the observer must orient the drawing table, with the x-axis being parallel to the Earths equator if projected onto the sun.

Next a Stonyhurst disk and its holder are placed on alignment holes recessed into the drawing table to keep the holder square to the edges of the table. The Stonyhurst disk is a circular piece of cardboard the same diameter as the solar image, which is equipped with lines of solar latitude and longitude, so as sunspot position can be calculated. Eighteen different Stonyhurst disks are used throughout the year to accommodate the varying orbital distances from the sun. The plane of the Stonyhurst is close to the focal length of the telescope, sunspots will be projected onto the disk, and so their heliocentric positions can be measured. These independent records of the sunspot latitudes have proved a good appendage for the Greenwich sunspot latitude records. As can be expected the use of drawings does lead to systematic differences compared to observatories where photographic plates were used [*Baranyi et al.*, 2001]. Also different sites have different scaling practises, making intercalibration necessary, this will be discussed in further detail in chapter 5.

Unfortunately the Mt. Wilson observatory only measured sunspot latitude, and for measurements of sunspot area and longitude, the USAF sunspot records must be used. The observations are made using the Solar Observational Network (SOON), from a variety of solar observatories around the United States, with the Mt. Wilson

observatory being the principle station, and other stations being utilised if observing is obscured at Mt. Wilson. The images are digitised and the areas of the sunspots are calculated from images on 16-inch photographic plates. Of late there has been some degree of confusion as to when the USAF sunspot measurements commenced, the NASA/NOAA web page (www.nasa.gov) states that the USAF sunspot area measurements were taken for the same period as the Mt. Wilson data, whereas the WDC (www.wdc.rl.ac.uk) and the Mt. Wilson observatory both state that the USAF sunspot area measurements commenced in 1981. As a result, the USAF sunspot area measurements can only be used with any confidence after 1981. Again there are some discrepancies present between the Greenwich observations and the USAF and these shall be outlines in Chapter 5.

Since there are concerns about the Greenwich and Mt. Wilson/USAF sunspot records for the period 1977-1981, another data set was sought to try and bridge this gap and thus create a continuous data set of sunspot observations that I would have confidence in. To this end I turned my attention to the USSR sunspot observations, which cover the period 1968-1991, and thus there is an overlap of a solar cycle in both the Greenwich and Mt. Wilson data sets. These "USSR" data were recorded using white-light photographic plates and hand drawn solar maps at a variety of observatories in the former Soviet Union. The most important observatory is the Pulkovo Main Astronomical Observatory near St Petersburg. The sunspot observations were originally published in the Solnechnie Dannie Bulletins, Monthly Reports, issued by the Main Astrophysical Observatory in Nauka, Leningrad under the auspices of the Russian Academy. They were also published in Catalogue Solnechnoi Deiatel'nosti, and can now be accessed from NASA/NOAAs National Geophysical Date Centre (www.nasa.gov). The intercalibration of the Greenwich, USAF/Mt. Wilson and USSR data sets are investigated in chapter 5.

Chapter 3: Investigation of the long-term changes in the solar photosphere and changes in the coronal source flux.

3. 1. Chapter overview.

Daily records of sunspot activity were first recorded at the Greenwich observatory on the 5th of May 1874, and reliable records were kept until December 1981, where upon the observatory at Mt Wilson continued the record until the present day. As will be discussed in Chapter 5, the intercalibration of these two data sets is not straightforward. These records contained measurements such as individual sunspot group latitude, longitude, number and area, and have proved extremely useful. Because of the lack of other reliable long-term datasets, these records have attracted interest as one of the few consistent records that are available for investigation of solar variability. In recent years, there has been a growing interest in century-scale variations of the Sun, with their possible implications for the terrestrial climate.

The coronal source surface is defined to be at a heliocentric distance of 2.5 R_s, and is where the solar magnetic field becomes approximately radial. Magnetic flux emerges through the solar photosphere in active regions and then rises up through the corona, as the opposite polarity footprints separate due to differential rotational, supergranular diffusion and meridional circulation. A small fraction of the flux that threads the photosphere (less than 10%) emerges through the coronal source surface [*Wang et al.*, 2000a], and is then classed as 'open' solar flux (as opposed to 'closed' flux, which loops back to the solar surface without threading the source surface). *Lockwood et al.* [1999a] have developed a method for estimating the annual means of the open flux, from observations geomagnetic activity, and this method can be applied to a data series, which exists from 1868 to the present day. The most striking feature of the coronal source flux data, is the long-term drift found to have occurred over the past 100 years, with the coronal source flux having doubled. This long-term drift is

superposed upon the 11-year solar cycle variation and has been noted in a variety of related data series. These include abundances of the cosmogenic ^{10}Be isotope [*Beer et al.,* 1988., *Lockwood,* 2001], and the occurrence of low-latitude aurora [*Pulkkinen et al.,* 2000]. It has been suggested that the upward drift in open flux derived by *Lockwood et al.*, [1999a] arises only from inaccurate inter-calibration of the geomagnetic index [*Svalgaard et al.,* 2003]; however the index is generated using data from both the northern and southern hemisphere, and the same trend is seen in the two separate data series as in the combined series [*Lockwood,* 2003]. This, along with the high correlations with independent but reliable data, virtually eliminates the possibility of calibration errors in the data.

Several solar parameters have been found to show similar long-term variations, such as peak sunspot number, sunspot area, the average heliographic latitudes of sunspots [*Pulkkinnen et al.,* 1999], and the length of the solar cycle [*Lockwood,* 2001]. *Solanki et al.* [2000] used a simple model of the continuity of open flux to show that the coronal source flux variation derived by *Lockwood et al.,* [1999a] was consistent with variations in the rate of emergence of flux through the photosphere, as quantified from the observed sunspot activity. Recently, *Lean et al.,* [2002] have used a sophisticated model of flux tube evolution (under the influence of diffusion, meridional flow and differential rotation) to arrive at the same conclusion.

As discussed in the introduction, *Lockwood and Stamper* [1999] found a correlation between the coronal source flux and total solar irradiance, and they found that using a simple linear regression they obtained a reconstruction of irradiance, that was similar to that proposed by *Lean* [2000b]. *Lockwood* [2001] has shown that on decadal time scales the correlation was very strong and statistically significant. The similarity of the TSI reconstructions to the *Lockwood and Stamper* extrapolation based on open flux, strongly implies that this relationship also applies on century timescales. Such a link could provide an explanation of the excellent correlations of paleoclimate records and cosmogenic isotopes discussed in the introduction. It should be stressed here that the good correlation between TSI and the open flux reported by *Lockwood and Stamper* [1999] and *Lockwood* [2002] on decadal timescales cannot automatically be applied to longer timescales because the physics of the connection is

not understood. In Chapter 6 of this thesis, I investigate the potential mechanisms of the connection in more detail.

Utilising the Greenwich and Mt. Wilson sunspots records available through the World Data Centre system, I here investigate variations in sunspot latitudes on both monthly and annual timescales. This strengthens and clarifies the association between surface magnetism and the open solar flux. An investigation is made between the coronal source flux, the spread of the latitudes, and the emergence rate of surface flux is made. The use of spread as a proxy for solar irradiance variations is also investigated.

3. 2. Annual means and standard deviations of sunspot latitude.

Daily values of sunspot group latitude are here employed for the period 1874-1999, and for each calendar year, average latitude and its standard deviation were calculated. As can be seen in figure (3.1) and figure (3.2), sunspot latitude in both the northern hemisphere and the southern hemisphere has varied considerably over the past 100 years with peaks in sunspot latitude being recorded in the year 1955 in the northern hemisphere, and in the year 1934 in the southern hemisphere. As can be seen in the figures, sunspot latitude is smallest at sunspot minima, and then increases rapidly as new cycle spots appear at higher latitudes closer to the poles, with the largest average latitude occurring around 2 years after sunspot minima. This is usually when all lower latitude spots have disappeared and only new spots appear on the solar surface. As the sun moves through its 11-year cycle, the average latitude slowly deceases as lower latitude spots appear and high latitude spots appear less frequently. This is consistent with the famous "butterfly diagram", first noted by Maunder in 1906, and as shown in figure 1.2. The intercalibration of the Greenwich and Mt. Wilson data is discussed in chapter 5.

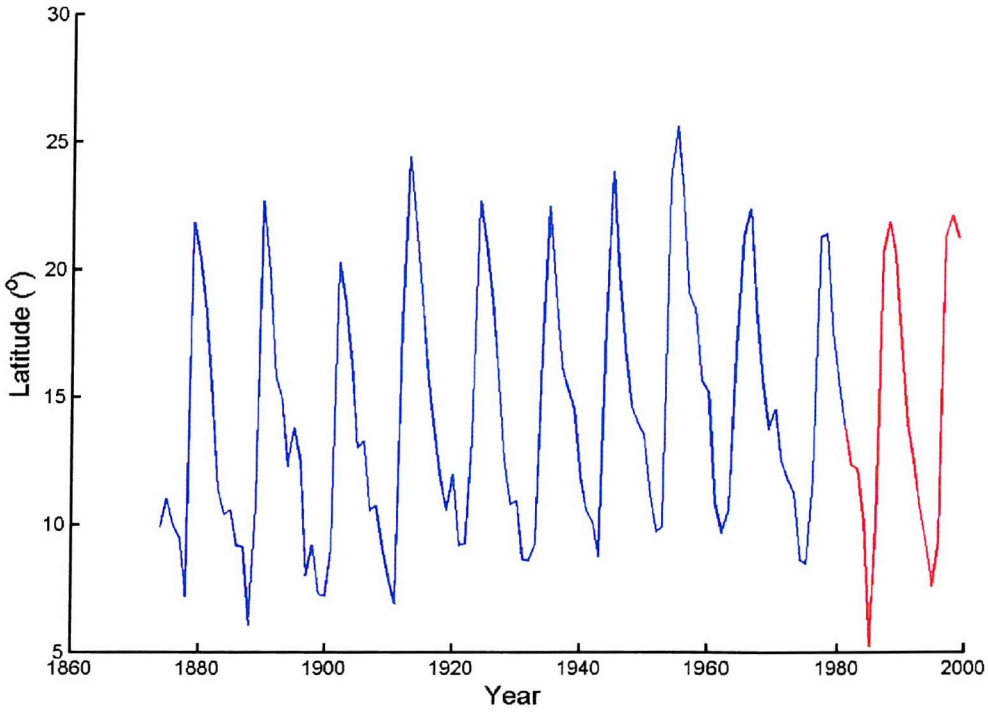


Figure 3.1. Composite of the annual means of northern hemisphere sunspot group latitudes as observed at Greenwich (blue) and Mt. Wilson (red), between 1874 and 1999.

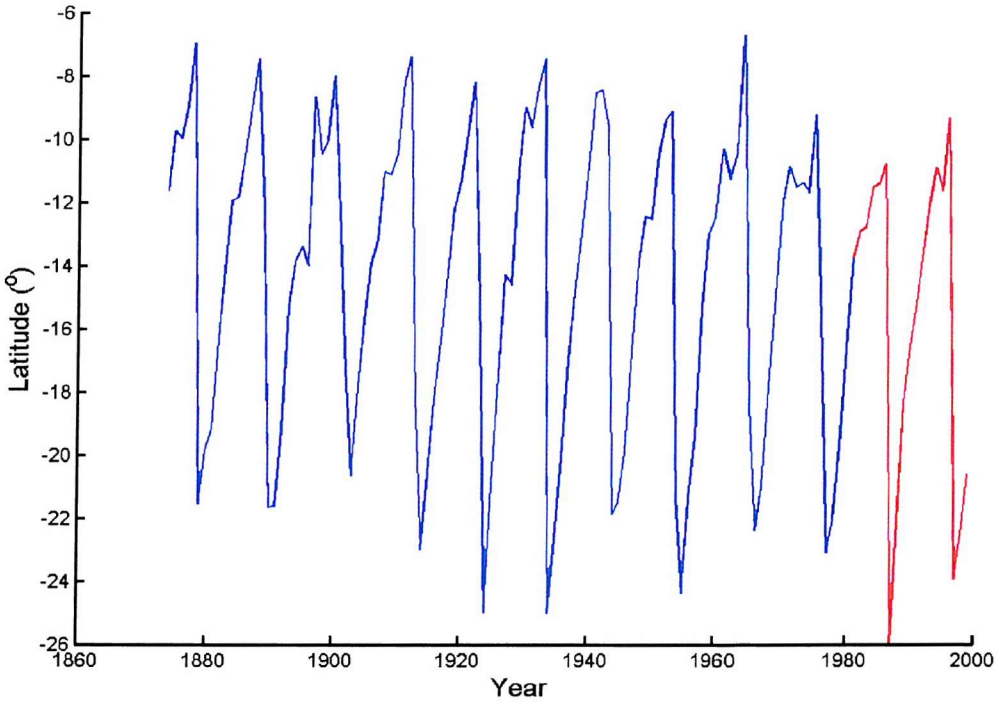


Figure 3.2. Composite of the annual means of southern hemisphere sunspot group latitudes as observed at Greenwich (blue) and Mt. Wilson (red), between 1874 and 1999.

Figures 3.3 and 3.4 shows annual sunspot 'spread' (the standard deviation of latitudes) between 1874-1981, for the northern and southern hemispheres respectively. It can be seen that sunspot spread has also increased over the period of the RGO observations. The most striking feature in the plots are the large 'spikes' in sunspot spread which occur just after sunspot minima, although it was quickly realised that these were associated with the overlap of the effects of the old and the new solar cycles.

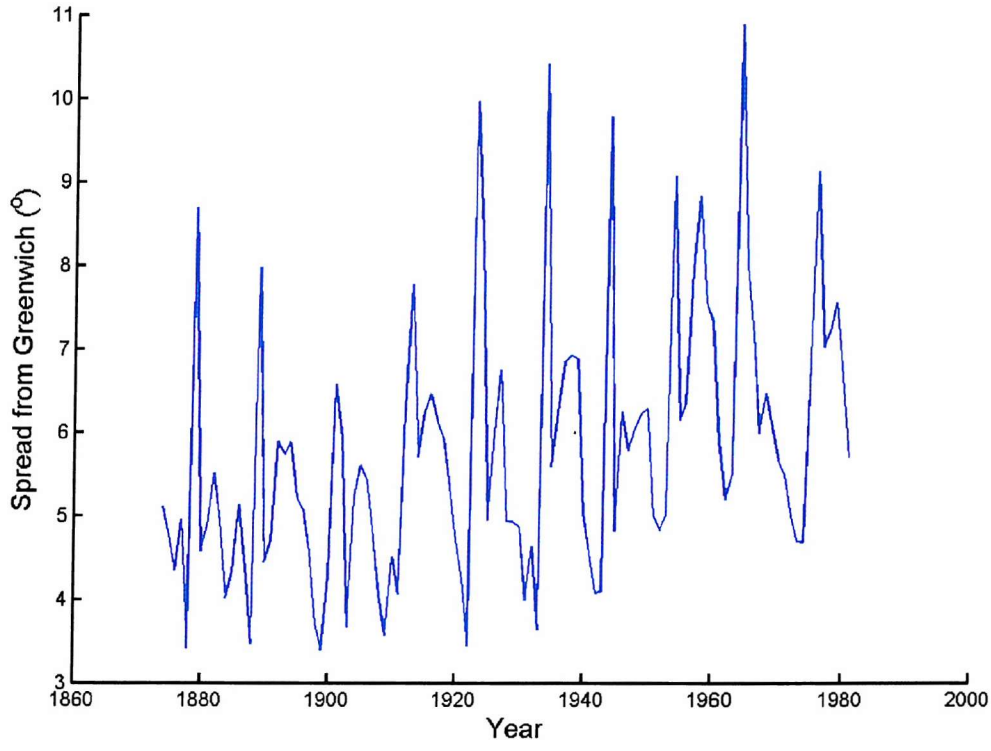


Figure 3.3. Annual standard deviation of northern hemisphere sunspot group latitudes as observed at Greenwich, between 1874 and 1981.

In the first few years after sunspot minima, both high (new cycle) and low (old cycle) spots exist on the solar surface, and these are in separate latitude bands. This overlap is not straightforward because low latitude spots from previous cycles can continue to appear for a number of years into the new cycle, whereas for other cycles the overlap is short-lived or may even be entirely absent. During the overlap, the spread is dominated by the fact that two bands exist, whereas at all other times it is determined by the widths of the lone latitude band. Hence, the spikes appear in figures 3.3 and 3.4.

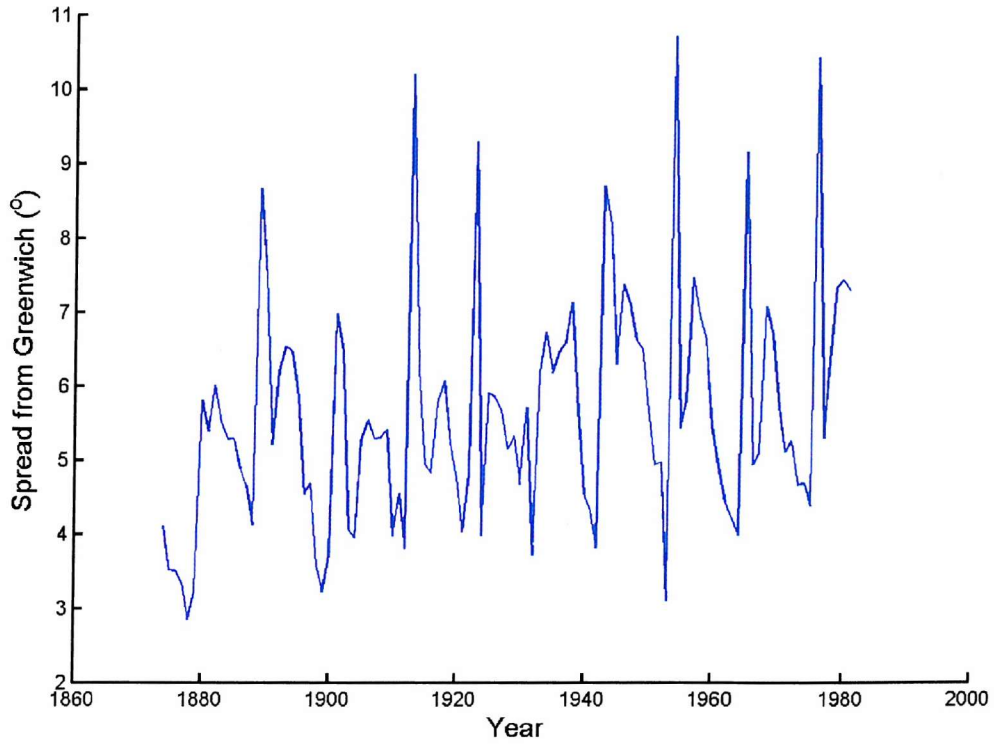


Figure 3.4. Annual standard deviation of southern hemisphere sunspot group latitudes as observed at Greenwich, between 1874 and 1981.

In order to generate a homogeneous dataset, in which the spread is an indicator only of the width of the latitude band, I separated the bands in which the sunspots appeared. To do this I identified a clear latitude at which no sunspots occurred during the overlap intervals. This "separator latitude" separated the new and old cycle sunspot bands. Using this separator latitude, I evaluated the spread in each band independently (and in each hemisphere). This can be seen for a sample overlap period in figure 3.5. As can be seen in the table below the separator latitude varies from cycle to cycle and is different in each hemisphere.

Year	*Northern separator latitude*	*Southern separator latitude*
1879	+15°	-15°
1889	+15°	-15°
1890	+15°	-15°
1901	+15°	-15°
1902	+12°	-15°
1913	+10°	-10°
1914	+8°	-10°
1923	+20°	-10°
1924	+12°	-15°

1933	+20°	-20°
1934	+15°	-10°
1943	+30°	-20°
1944	+15°	-15°
1954	+18°	-18°
1963	+20°	-20°
1964	+18°	-20°
1965	+15°	-15°
1975	+11°	-15°
1976	+20°	-20°

Table 3. 1. Latitude band separators at solar minimum for Greenwich data.

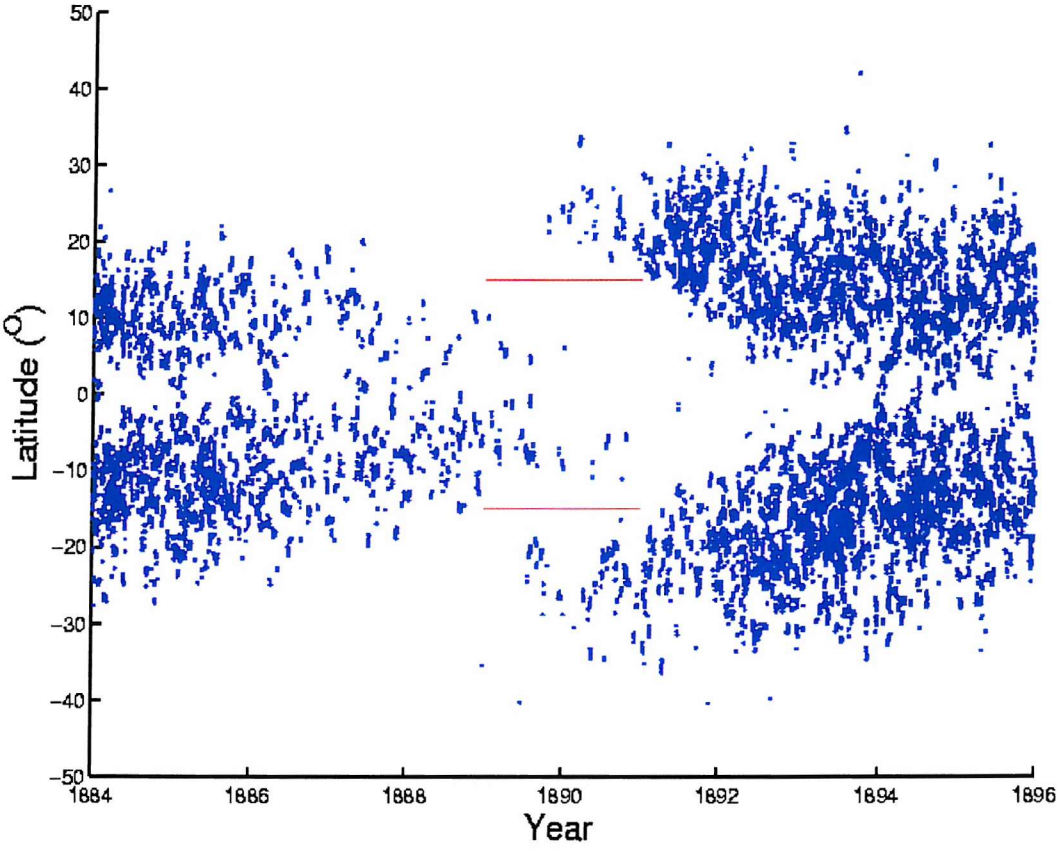


Figure 3.5. Sunspot butterfly plot for the period 1884-1896. Each blue spot marks a sunspot group observation. The overlap between the new and old cycle spots can be seen in years 1889 and 1890. The red line denotes the separator latitude.

In order to preserve information from both bands in the overlap period, I evaluated the spread of the two bands and then combined the separate spread values in each hemisphere as the square root of the sum of the squares. The results of this can be seen in figures 3.6 and 3.7. As can be seen the spikes in the spread data have been removed and so I have a true representation of how the width of the latitude bands varied over the duration of the Greenwich observations. The solar cycle variations are

similar to those found in sunspot number, with the peaks in spread occurring around solar maximum, and the smallest spread occurring around solar minimum, although unlike in sunspot number the spread does not fall to relatively constant and small values at every minimum. What is most interesting is the increase in spread over the interval of the Greenwich observations. Spread is found to increase slightly in the first two solar cycles then drop slightly at the turn of century, then again at around 1960. Differences can be seen in each hemisphere, with the drop at the turn of the century being less evident in the northern hemisphere, although the overall long-term drift can be seen in both hemispheres.

Year	*Northern separator latitude*	*Southern separator latitude*
1963	+20°	-20°
1964	+18°	-20°
1965	+15°	-15°
1975	+11°	-15°
1976	+20°	-20°
1986	+21°	-13°
1987	+10°	-10°
1996	+20°	-21°
1997	+10°	-15°

Table 3. 2. Latitude band separators at solar minimum for Mt. Wilson data.

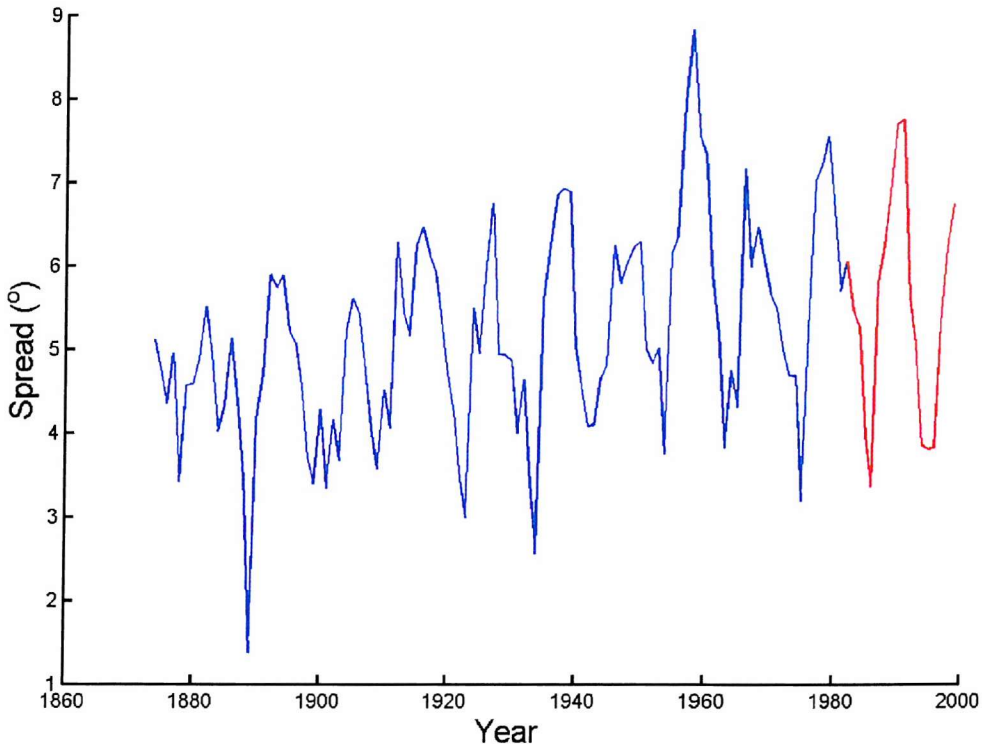


Figure 3.6. Composite of the annual combined spread of northern hemisphere sunspot groups latitudes, σ_N, as observed at Greenwich (blue) and Mt. Wilson (red), for the period 1874-1999.

To continue the spread data set, I again used the Mt. Wilson observations, using the same procedure to deal with the overlap periods. The separator latitudes defined are given in Table 3. 2. As with sunspot latitude, the overlap period between the two data sets was used to check the validity of continuing the spread measurements using the Mt. Wilson observatory, and this will be investigated in further detail in Chapter 5. A continuous data set was created, and the results can be seen in figures 3.6 and 3.7 show annual sunspot spread from 1874 –1999 for the northern and southern hemispheres respectively.

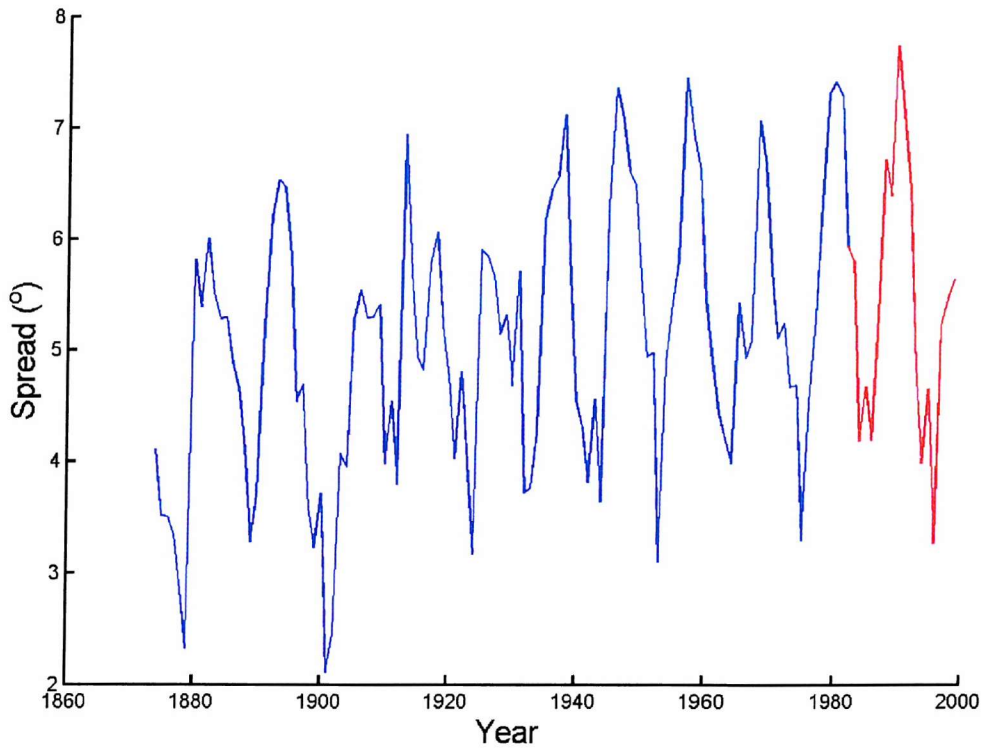


Figure 3.7. Composite of the annual combined spread of southern hemisphere sunspot groups latitudes, σ_S, as observed at Greenwich (blue) and Mt. Wilson (red), for the period 1874-1999.

To get a better idea of how the spread has varied over the duration of the combined data set I combined the northern hemisphere and southern hemisphere values (σ_N and σ_S, respectively) into one parameter, $(\sigma_N + \sigma_S)/2$ (figure 3.8). Averaging the spread in both hemispheres removes variations that are localised to a single hemisphere, and gives a whole Sun estimate that can be compared with other global solar parameters.

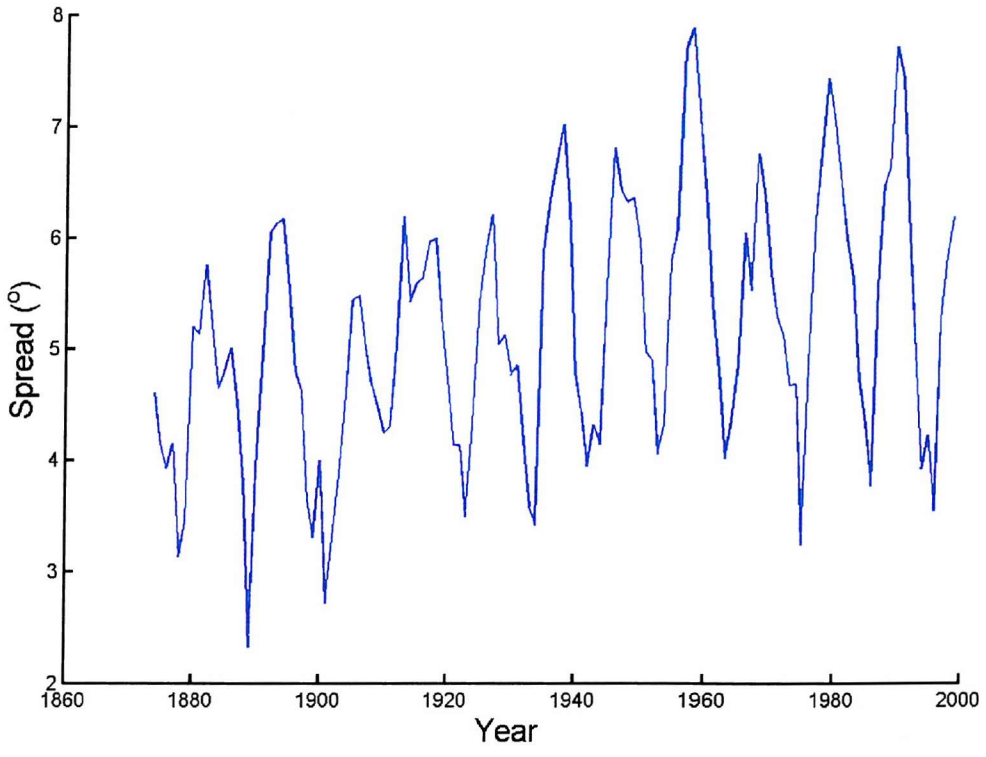


Figure 3.8. The average of the northern and southern hemisphere sunspot group latitude spread, $(\sigma_N + \sigma_S)/2$, for the period 1874-1999.

3. 3 Annual coronal source flux and the flux emergence rate

Figure 3.9 shows the variations of annual means of the coronal source flux (F_S) as derived from the geomagnetic aa index by *Lockwood et al.* [1999a] and compares it with the variation in $(\sigma_N + \sigma_S)/2$. The two data sets show a great many similarities, in their averages as well as their maxima and minima. Both show similar solar cycle oscillations with peaks occurring at solar maximum, superposed upon a long-term increase. F_S also shows brief decreases around the turn of the century and around 1960, which were also found in the spot latitude spread data. Although there are a great many similarities between the overall trends in the data sets, the ratio of the solar cycle amplitude to the magnitude of the long-term drift are different. For the coronal source flux, F_S, the amplitude of the variation during the period of solar cycle 22 (1986 –1995) is 3.1×10^{14} Wb, whereas the average for the whole of cycle 22 is larger than that for cycle 14 (1901-1913) by 2.91×10^{14} Wb. The corresponding values for $(\sigma_N + \sigma_S)/2$ are 3.95° and 1.11°. Therefore the ratio of the long-term drift (since

1900) to the recent solar amplitude is roughly one for the coronal source flux, but only about 0.3 for $(\sigma_N + \sigma_S)/2$. This difference gives rise to the scatter in figure 3.10, nevertheless a correlation is found between the two data sets of 0.66 for simultaneous data, which increases to a peak of 0.81 if the coronal source flux lags behind the spread by the optimum delay of 1 year. This correlation is significant at greater than the 98% level.

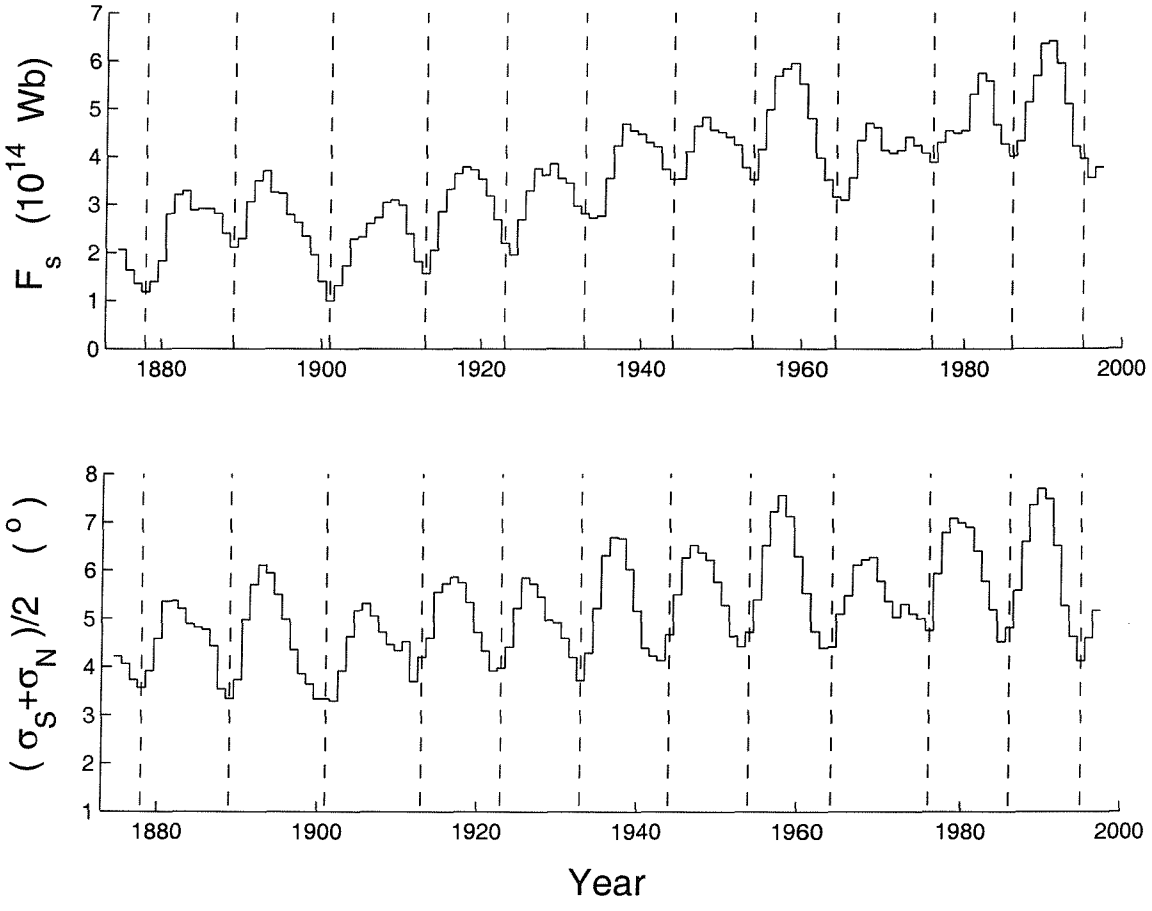


Figure 3.9. (a) The variation of annual means of the coronal source flux, F_S, as estimated from geomagnetic activity using the method of *Lockwood et al.* [1999a]. (b) The average of the standard deviations of sunspot group latitudes, $(\sigma_N+\sigma_S)/2$. Vertical dashed lines mark sunspot minima.

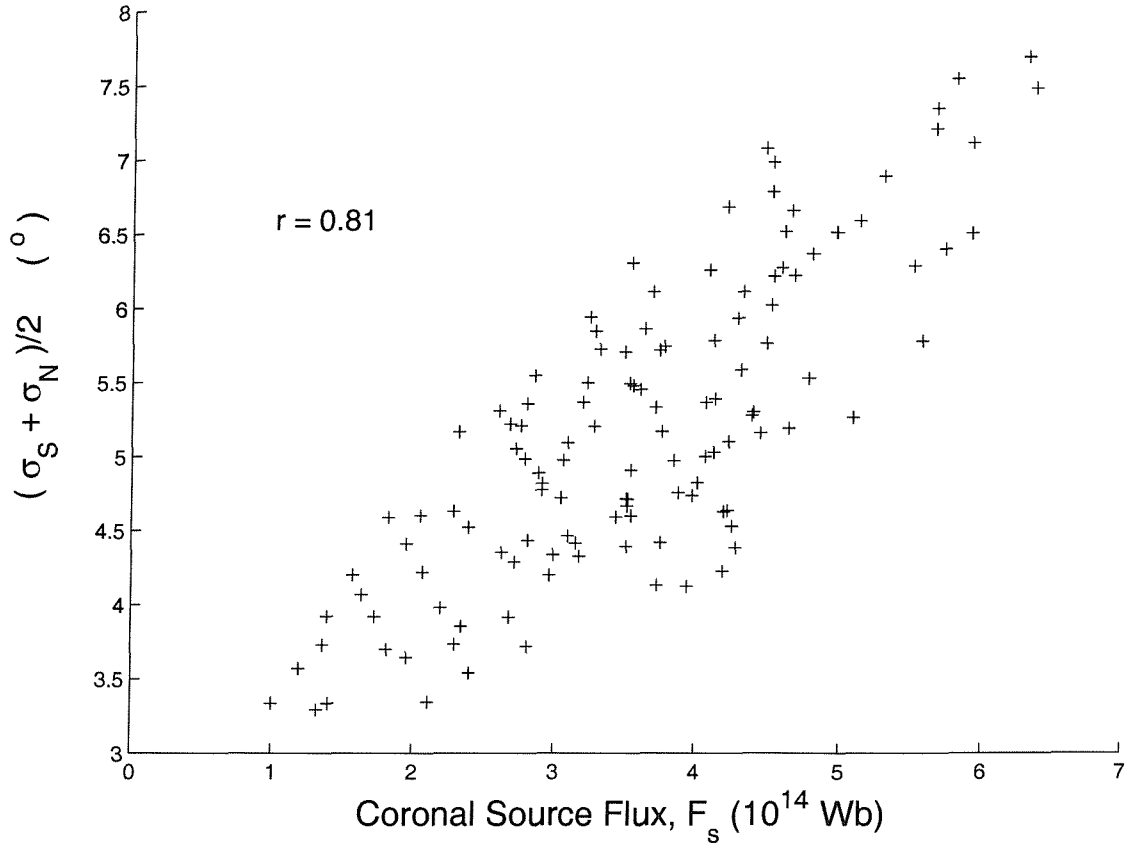


Figure 3.10. Scatter plot of the annual means of the coronal source flux, F_s, and the combined northern and southern sunspot latitude spread $(\sigma_N+\sigma_S)/2$.

Solanki et al. [2000], have obtained an excellent match to the variation of F_s derived by *Lockwood et al.* [1999a], using a simple continuity model. The key equation in their model is that of the continuity of open flux:

$$\frac{dF_s}{dt} = E - \frac{F_s}{T} \tag{3.1}$$

This can be rewritten to define the variation in the flux emergence rate:

$$E = \frac{dF_s}{dt} + \frac{F_s}{T} \tag{3.2}$$

where T is the time constant for the destruction of open flux and E is the rate of its emergence through the coronal source surface. Because the model does not consider the distribution of open flux over the solar surface, T is an effective average for all open flux. *Solanki et al.* equate E to γE_p, where E_p is the rate of flux emergence

through the photospheric active regions and $\gamma = (1+ \tau_t/\tau_a)^{-1}$, and so depends on the time constants for flux annihilation in active regions (τ_a), and for transfer of flux from active regions to the network (τ_t). *Solanki et al.* estimate γ to be 0.015 and derive a best fit with a time constant T close to 4 years. Because F_s and dF_s/dt is known, I can use equation (3.2) to compute the flux emergence rate E for an assumed value of T. *Solanki et al.* estimate E_p from R (sunspot number), but need to allow for the fact that only a small part of the flux emerges from sunspots (total area A_s), and a much larger part emerges as small flux tubes from active region faculae (total area A_f). Thus they multiply R by an area factor (which varies over the solar cycle) to give E_p proportional to $(1 + A_f/A_s)R$. *Solanki et al.* then use an empirical relationship for (A_f/A_s) that allows E_p to be computed from R. Sunspot spread $(\sigma_N+\sigma_S)/2$ is a more direct indicator of the area of active regions. It differs from the area of sunspot groups because it includes the area between the groups.

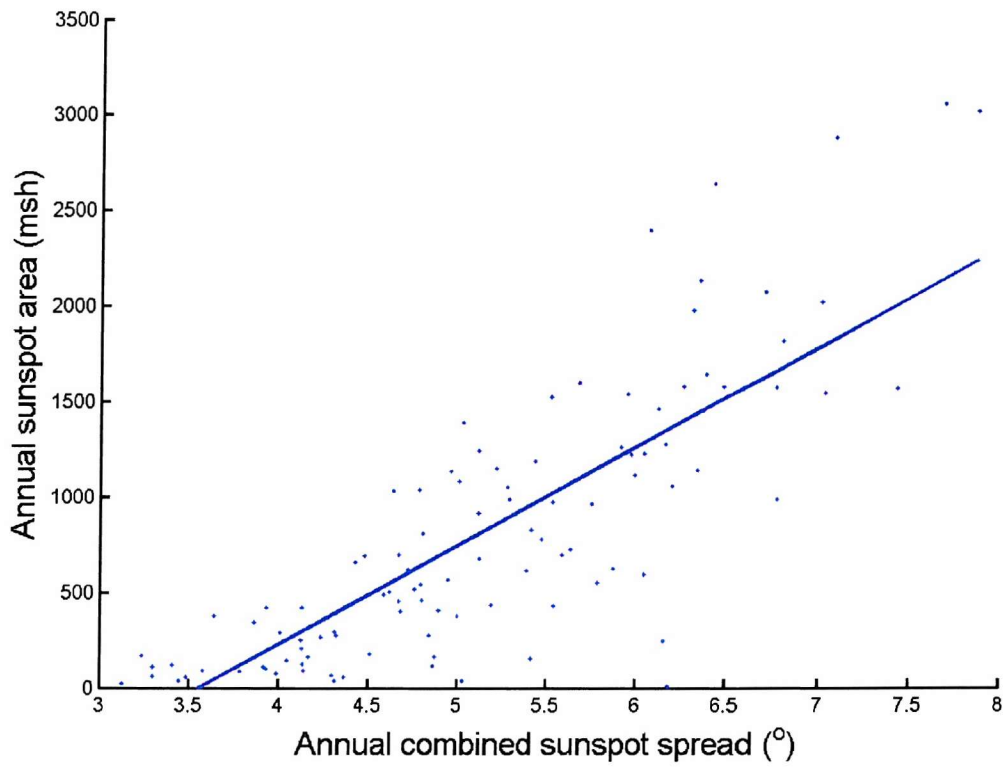


Figure 3.11. Scatter plot of standard deviation of sunspot group latitude and sunspot group area (in millionths of a solar hemisphere).

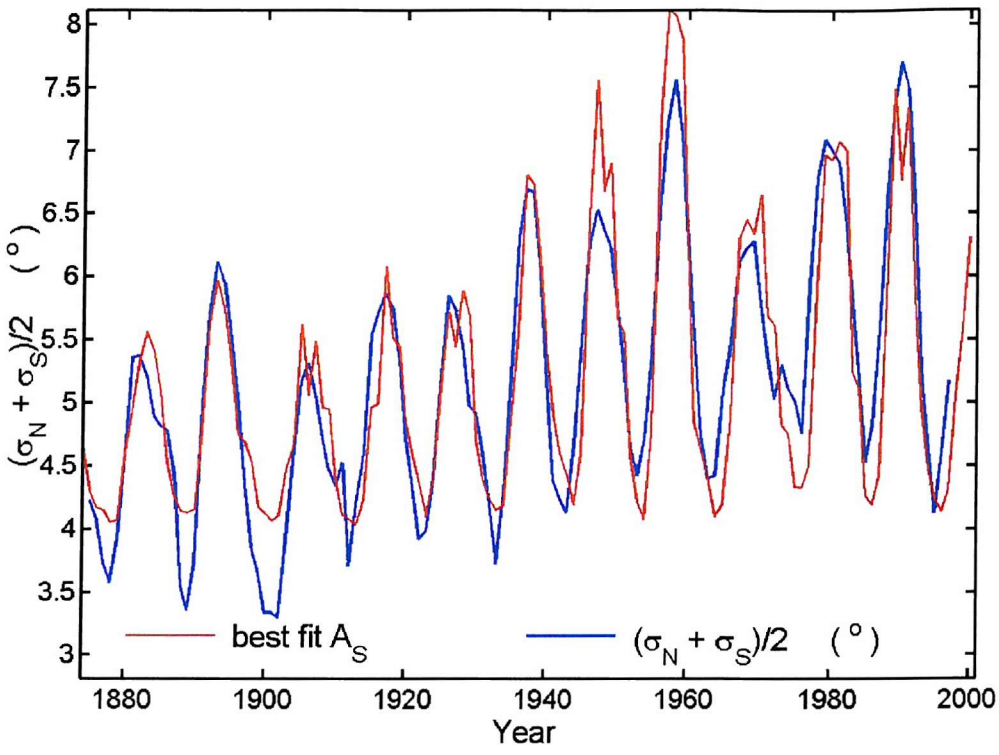


Figure 3.12. Annual Sunspot latitude spread from observations, and reconstructed using sunspot area measurements, for the period 1874-2000.

This difference can be seen in figure 3.11, which shows the best linear regression fit of the sunspot group area, A_s, with the sunspot spread $(\sigma_N + \sigma_S)/2$. The correlation between these two parameters is r = 0.89, which is significant at greater than the 99.99% level. However, figure 3.12 reveals that although the agreement is very good at sunspot maximum, there are important differences at sunspot minimum. Specifically the sunspot area (like sunspot number) returns to near-zero values at each minimum (although a slight upward drift in minimum values of both R and A_s did occur during the twentieth century), whereas there is considerable drift in the minima of $(\sigma_N+\sigma_S)/2$ which mirrors that in the open flux, F_s. Figure 3.12 shows that the few active regions that do exist at solar minimum are more spread out into the recent past than they were around 1900. When considering open flux emergence, this may be a significant factor.

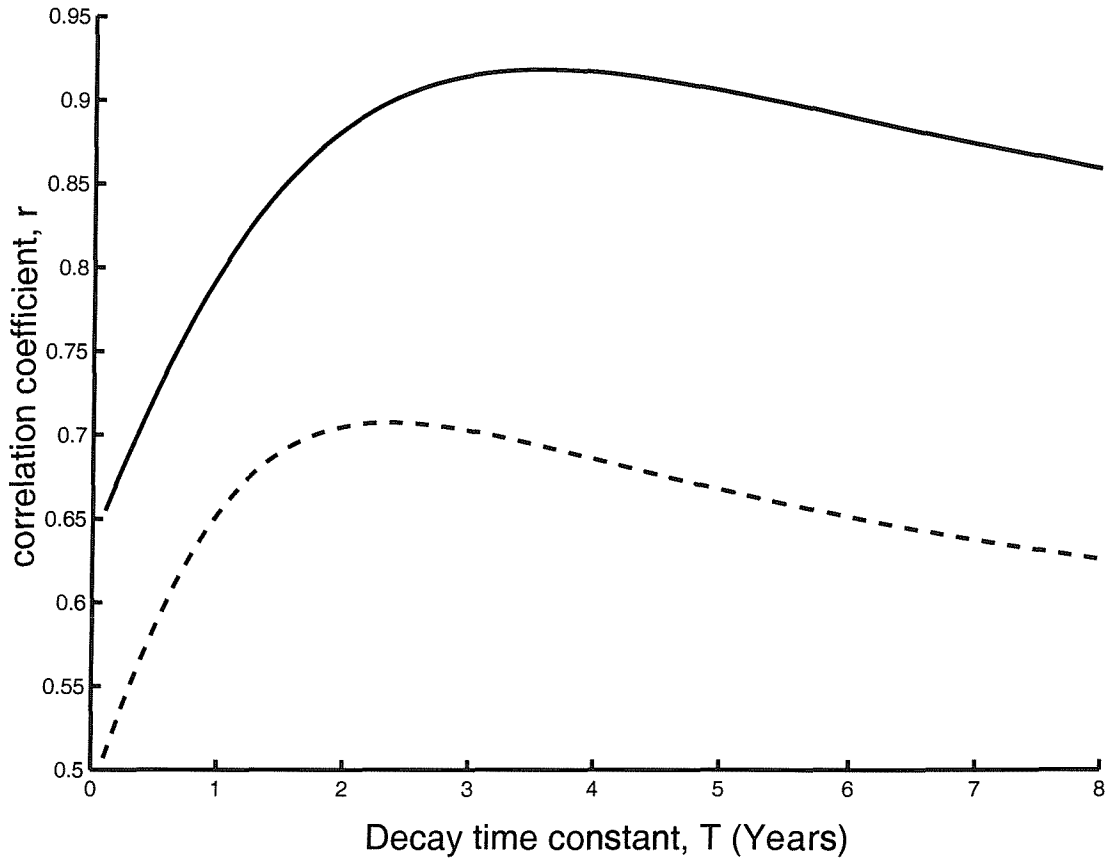


Figure 3.13. The correlation coefficient between the flux emergence rate $E = \{dF_s/dt + F_s/T\}$ and the mean of the standard deviations of sunspot group latitudes $(\sigma_N+\sigma_S)/2$, as a function of the flux decay time constant, T. The dashed line is for unsmoothed data, the solid line for 3-year running means.

Wang and Sheeley [2002], *Wang et al.* [2002] and *Mackay* et al. [2002] have produced models of the evolution of flux, which emerges in a single bipolar magnetic region (BMR), under the influence of differential photospheric rotation, meridional surface flow and diffusion. The flux emerging through the photosphere is partly destroyed by reconnection at the polarity inversion line (PIL) between the inward and outward surface flux which lengthens with time: the surviving flux rises through the corona and becomes open when it reaches the source surface. When realistic distributions of BMRs are included [*Mackay and Lockwood,* 2002; *Lean et al.,* 2002], reconnection takes place at PILs between different BMRs as well as at those at the centre of each BMR. The extent and magnetic shear across these extra PIL's will increase if the ratio $A_s/(\sigma_N+\sigma_S)$ increases. Thus the rate of emergence of surviving flux E_p might be expected to be proportional to A_s and to the ratio $(\sigma_N+\sigma_S)/A_s$. This suggests it is worthwhile considering the use of $(\sigma_N+\sigma_S)$ as an indicator of E. Figure

3.13, shows the correlation coefficient between $(\sigma_N+\sigma_S)/2$ and E, as a function of the flux decay time constant, T, used with equation (3.1) to derive E, from the observed dF_s/dt. The dashed line shows the results for annual means, but because relatively large fluctuations are introduced in taking the temporal gradient of F_s, three point running means give the higher correlations shown by the solid curve. The peak correlation is 0.92 at T =3.6 years for the smoothed data, which is close to the estimate of 4 years obtained by *Solanki et al.* Using a Fischer -Z test, I find the correlation coefficient is significantly different (at the 95% level) from its value at $T \leq$ 2.8 years and $T \geq 4.5$ years and this gives an uncertainty in the best T estimate as $\pm$ 0.8 year. The unsmoothed data gives $T = 2.6 \pm 1.2$ years which although is lower than the smoothed data, it is still consistent within the uncertainties. Figure 3.14 shows the scatter plot of the three year running means of E and $(\sigma_N+\sigma_S)/2$ for the optimum decay time constant of T =3.6 years, as defined by figure 3.13. The solid line is the best-fit linear regression, and this fit is used to predict the variation of the flux emergence rate E from the observed $(\sigma_N+\sigma_S)/2$. The result is the solid line in figure 3.15, which compares very well with the variation computed from F_s (dot dashed line).

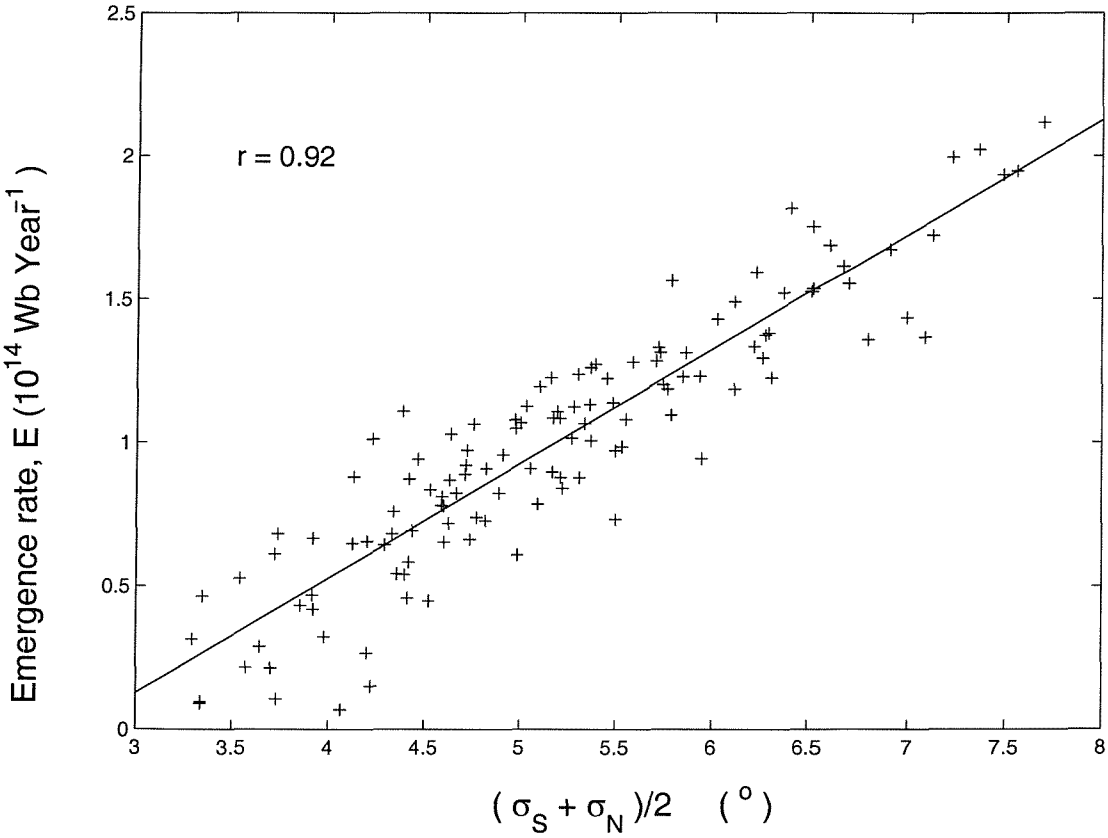


Figure 3.14. Scatter plot of the flux emergence rate E, and the standard deviation of sunspot group latitudes for the best-fit of T.

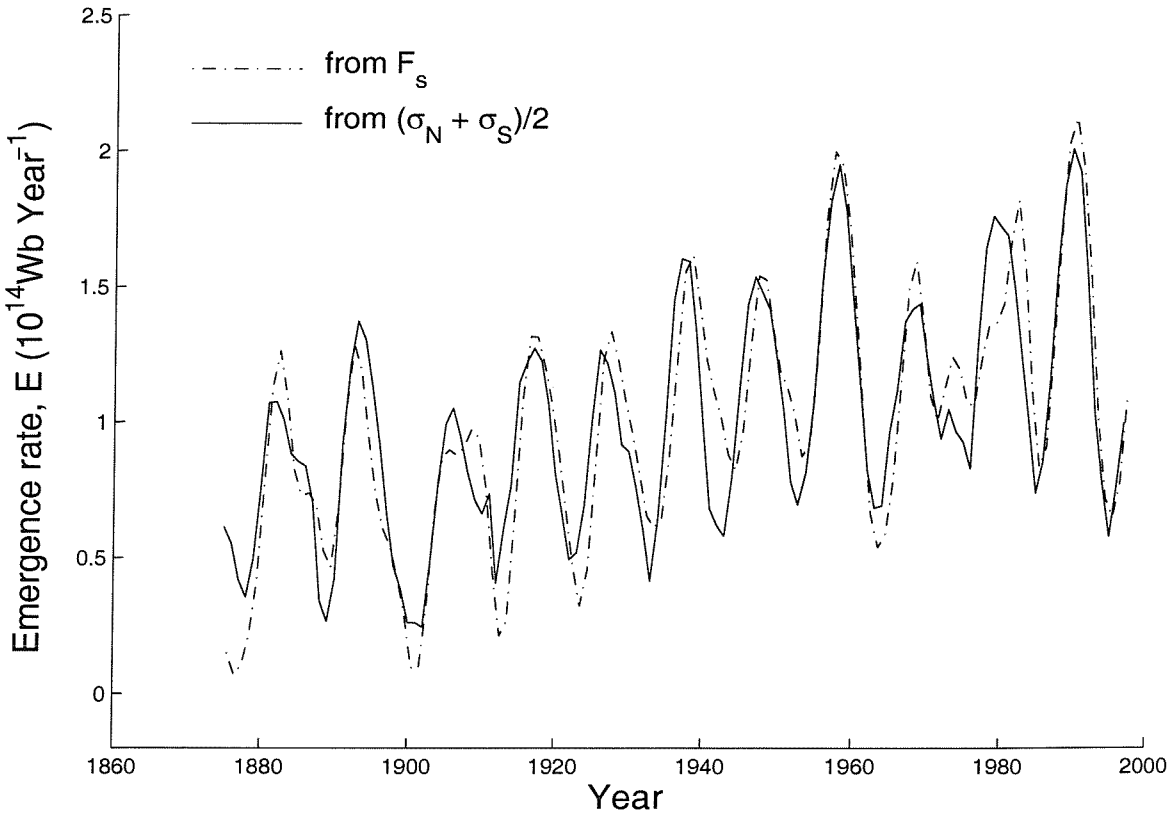


Figure 3.15. Variation of the flux emergence rate calculated from F_s, and from sunspot latitude spread.

As a test of these estimates of E, derived from the spread of sunspot latitudes, I can use them along with the best-fit T of 3.6 years, as an input into equation (3.1) to model the variation of F_s, the results are shown in figure 3.16. The model predictions (solid line) are insensitive to T over its most probable range (2.6-4.5 years) and are a very close match to the values observed by *Lockwood et al.* (dot-dash line). As can be seen the solar cycle variation and the long-term drift are reproduced, showing that sunspot latitudinal spread is indeed a good indicator of E_p, on annual timescales. This work confirms a relationship between open flux accumulation and decay, and active region features on the solar surface. Although the simple global continuity employed by *Solanki et al.* and inherit in equation (3.1) is undoubtedly simplistic, fuller modelling of emerged flux evolution confirms such a relationship [*Lean et al.* 2002].

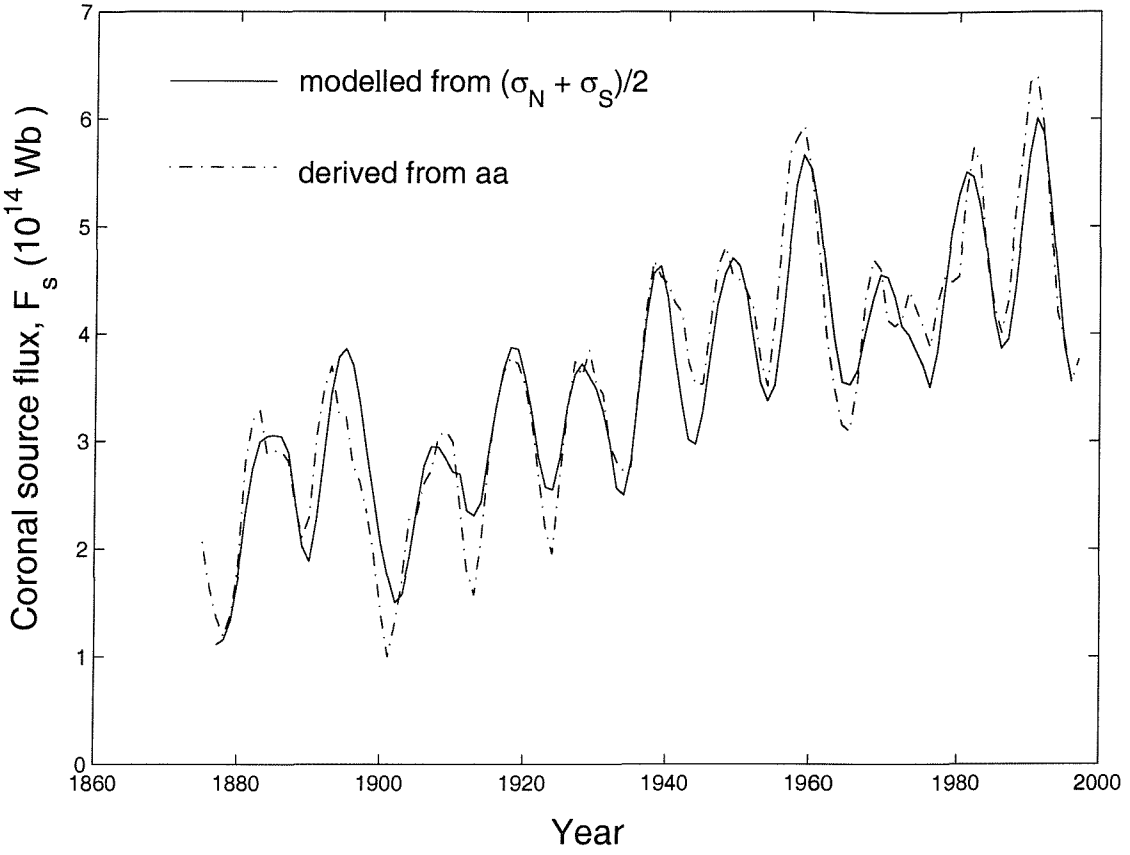


Figure 3.16. The variation of the coronal source flux, F_S, and as modelled using the best fit of E, for the best-fit flux decay time constant of 3.6 years.

3. 4. Monthly sunspot latitude and spread data

Monthly averages of sunspot latitude and spread were also calculated from the Greenwich and Mt. Wilson observations using the same procedure as applied to annual means in Chapter 5. A composite average of the northern and southern hemisphere average latitude can be seen in figure 3.17, and the full spread sequence at one-month resolution is shown in figure 3.18.

As expected the results are similar to the annual data, but with some notable differences. Again, it can be seen that sunspot latitudes have steadily increased since 1874 with maximums occurring in solar cycle 19 and 17 in the northern and southern hemispheres respectively. At early sunspot minima there are months where the latitude is zero, this is due to there being no sunspots visible to observers on the solar surface, consistent with being at relatively low activity. As the sun has become

increasingly magnetically active over the past 100 years sunspot numbers have increased, hence there are relatively few months when there are no sunspots visible on the solar surface. As seen in the annual data, the average latitude of sunspots at solar minimum has increased.

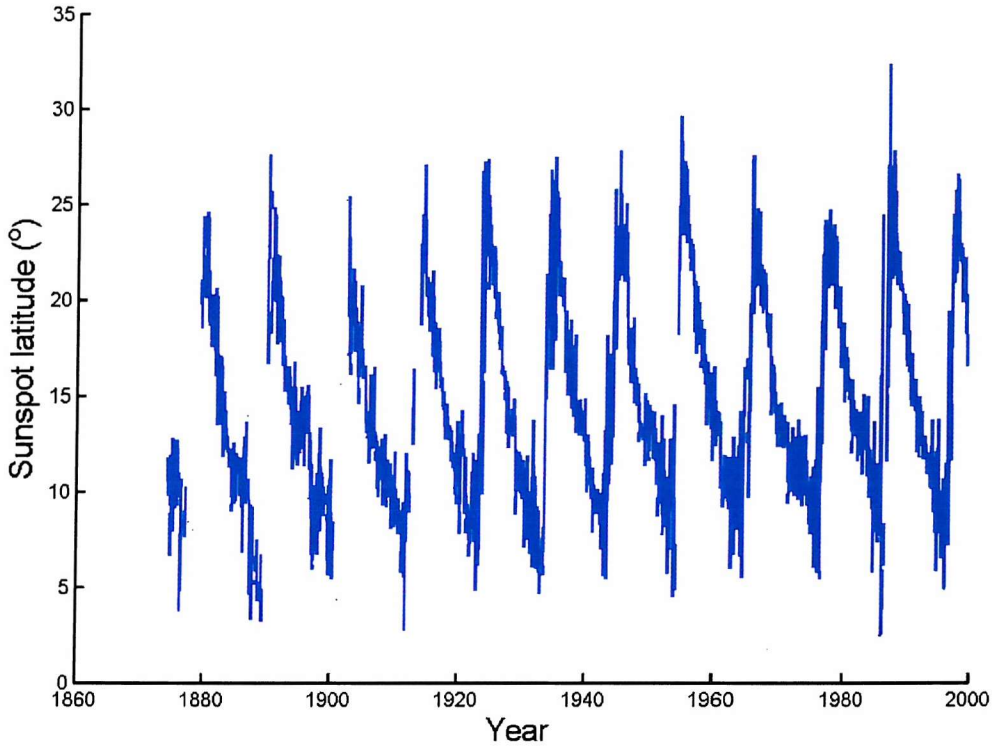


Figure 3.17. Monthly means of the variation of sunspot group latitudes.

The monthly spread data reveals there are usually only a couple of months at every solar minimum when the high and low-latitudes bands of spots co-exist. However as the sun has increased in activity over the past 100 years, the frequency of these overlap periods has increased, meaning that either the appearance of new high latitude sunspots is occurring earlier than usual, and/or that the old low latitude spots are persisting for a longer period than they did previously. Thus, the monthly spread data is treated in the same fashion as the annual data as outlined in section 3.2, and the corrected monthly data series can be seen in figure 3.18. As can be seen there are a number of months at each solar cycle that the spread drops to zero, and this occurs when there are no visible sunspots on the solar surface. The occurrence of these periods has been found to have decreased over the past 100 years, with the increase in sunspot activity. Again, it can be seen that the values at both maximum and minimum have increased, as was found in the annual data series.

The method of *Lockwood et al.,* [1999a] to derive the open solar flux, F_s, from the geomagnetic activity index aa requires the use of one year averages. There are three reasons for this:

1: There are seasonal variations in ionospheric conductivity, which alter the magnetic deflection seen at the two magnetometer stations used to derive the aa index.

2: There is an annual variation in the Sun-Earth distance caused by the non-circularity of the Earth's orbit which causes a very small, but regular annual cycle in the solar wind conditions around the Earth

3: The *Russell-McPherron* [1973] effect, in which solar wind magnetosphere coupling is modulated by the tilt of Earth's magnetic dipole axis with respect to the Sun-Earth line, causes a strong semi-annual variation.

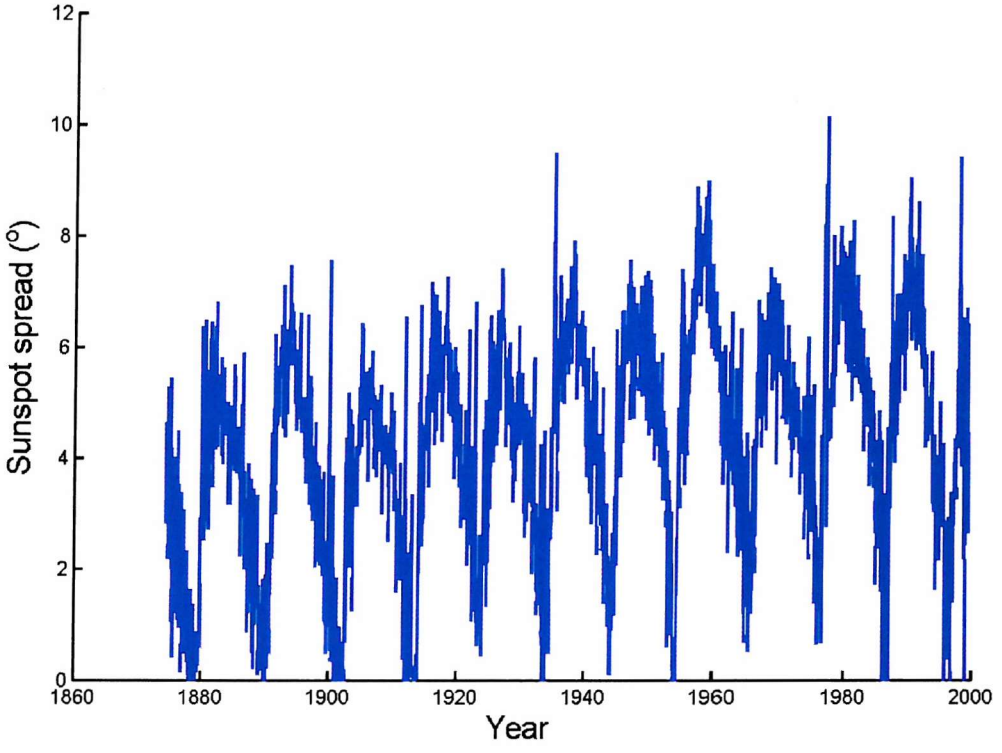


Figure 3.18. Monthly variations in the standard deviation of sunspot group latitudes.

I here average out all three effects by utilising a 12-month data window, which is progressed by a month at a time, generating monthly values of F_s for the period

1886-2001. However, it should be noted that only every twelfth data point is fully independent and that these are effectively a form of 12-month running means of monthly data, to which noise has been added by the procedure used to estimate F_s from the aa index data. This effective smoothing increases the auto-correlation of the data series at lag 1 and so reduces the significance of any correlations obtained (compared with what would be obtained for independent monthly F_s values as obtained, for example, from monthly means of the radial component of the interplanetary magnetic field). The monthly F_s values can be seen in figure 3.19

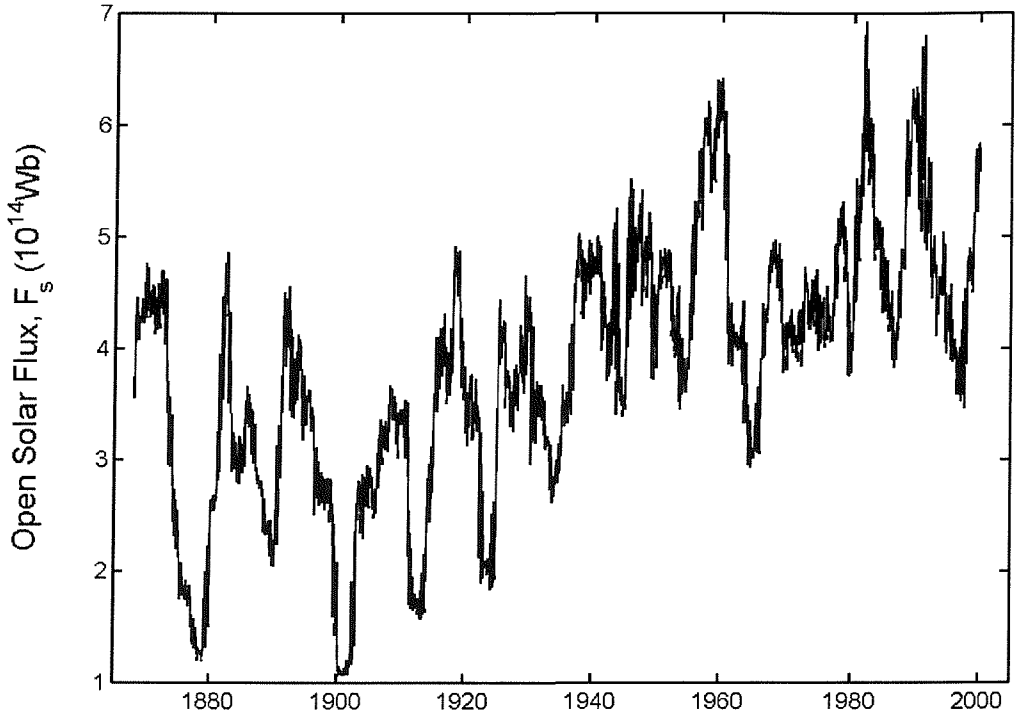


Figure 3.19. Monthly variation of the coronal source flux, F_s.

One of the most striking features of the monthly F_s data set is the strange variation of F_S through solar cycle 20, where there is little or no solar cycle variation evident, and F_s remains almost constant through the solar cycle. The reason for this is unknown. Solar cycle 20 followed the strongest solar cycle ever observed (the peak R of cycle 19 is the largest recorded) and was noted as something of an anomaly with several parameters showing a weaker and unusual variation in this cycle. The similarities between *F*s and sunspot spread are still noticeable in the monthly data sets of the two parameters, although the correlation is lower than for the annual data sets because of the larger noise fluctuations (r = 0.63). The significance is increased to

99% because the number of data points is 12 times larger (even though the F_s data are effectively 12-point running means).

Thus using monthly data, I obtain a more significant correlation than for the annual data, but general trends (solar cycle variation superposed on a long-term drift) are the same as in the annual data.

3. 5. Chapter discussion and summary.

The study presented here confirms the long-term increase in the latitudes of sunspots reported by *Pulkkinen* [1999], but also reports an increase in the area of the photosphere in which they can appear. This area is highly correlated with the area, A_S, of the Sun covered with sunspot groups at any one time. However, that correlation is largely due to data from solar maximum. At solar minimum the area A_s always returns to very small values, whereas the area of the latitude bands in which spots occur shows a long-term trend at solar minimum as well as at solar maximum. This latitude spread has been shown to correlate strongly with the coronal source flux, and the open flux emergence rate on both monthly and annual timescales. The sunspot spread data shows that the area of the region in which sunspots can occur, has increased steadily over the past 100 years, in a similar way to the coronal source flux.

Sunspots are the largest of the magnetic flux tubes, which thread the photosphere, smaller flux tubes being classed as micropores and faculae. Faculae occur in the immediate vicinity of sunspot groups (active region faculae) as well as over the entire surface near the joins of supergranulation cells (network faculae). There is also emerged flux in 'ephemeral' regions, which may in part be active region remnants. Smaller flux tubes, of weak strength make up the distribution seen in the quiet Sun. Thus where sunspot numbers quantify the flux emerging in large flux tubes by including the areas between sunspots groups, the sunspot spread could be regarded as an indicator of flux emergence in sunspots and the smaller flux tubes between them. This being the case, the long-term change in the latitudinal width of active regions bands would indicate an upward drift in the rate of emergence of flux in tubes of all sizes, through the solar surface, E_p.

However, the emergence rate E_p through the solar surface does not directly control the open solar flux and thus the strong and significant correlation between the open solar flux F_s and the latitudinal spread of active region bands, $\sigma = (\sigma_N+\sigma_S)/2$, may not be caused by a variation in E_p. Of the flux that has emerged through the photosphere only a small (and variable) fraction (typically a few percent) rises through the corona and so emerges through the coronal source surface to become 'open' solar flux. The remainder of the surface flux is destroyed by reconnection at polarity inversion lines or simply subsides back through the solar surface. Numerical models of the emerged flux evolution, under the influence of differential rotation, meridional flow and diffusion (e.g. *Mackay and Lockwood,* 2001; *Lean et al.*, 2002) show a huge variety of lifetimes of emerged and open flux. Open flux is embedded in the solar wind flow and so it does not subside back into the Sun and is only lost by reconnection at PILs (i.e. coronal current sheets). Some open flux is lost relatively quickly whereas some migrates all the way to the poles and collects in the polar coronal holes. This behaviour is reproduced in the modelling by *Lean et al.,* [2002], shown in figure 3.20, which reproduces the magnetograms presented in figure 1.3. Some of the polar open flux may remain for several years until open flux of the opposite polarity, emerging in the next solar cycle, begins to accumulate there. The simple model of *Solanki et al.* [2000] is concerned only with the continuity of the total amount of open flux and so uses an effective average of the loss time constant. The fact that this is surprisingly high must reflect long-lifetime flux in the polar coronal holes.

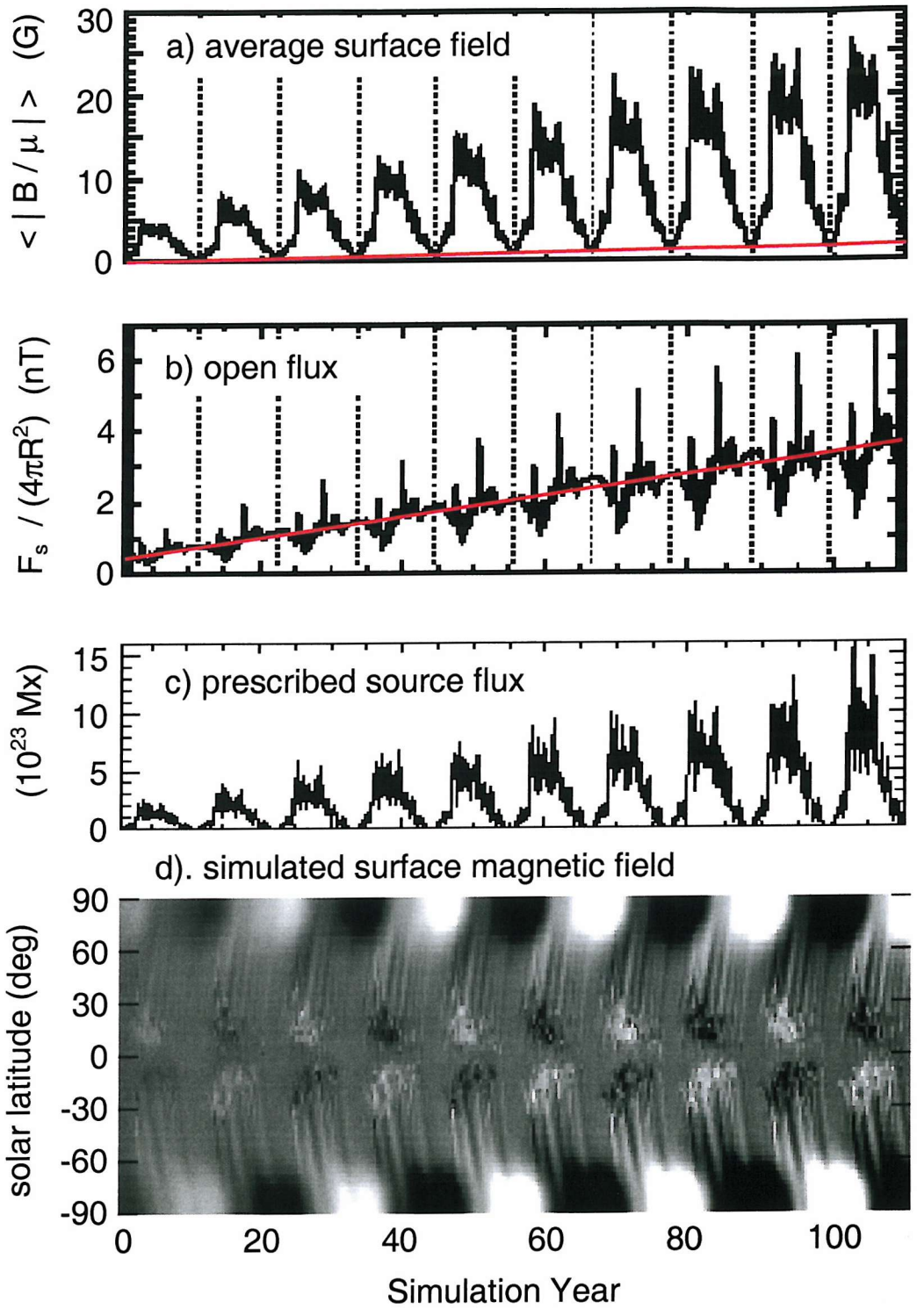


Figure 3.20. Simulation of open flux and surface flux by Lean et al., (2002). (a) The output total surface flux, expressed as $<|B / \mu|>$, (b) the output open flux $F_p/(4\pi R_s^2)$, (c) the input emerged surface flux, (d) the simulated latitude-date magnetogram for the central solar meridian.

I find that the spread of the active region bands σ is a good proxy for the open flux emergence rate *E* of open flux required by the *Solanki et al.* model. One reason for this is likely to be a good proxy for the emergence rate through the photosphere (in

sunspots, active region faculae and ephemeral regions), E_p. However, it may be also be that the fraction of emerged flux, which survives to become open decreases as σ is lower, because polarity inversion lines are more extensive and carry larger current. Thus, σ maybe a better proxy for open flux emergence rate than a complex function of an active region index (sunspot number, sunspot group area, active region facular area). Detailed analysis of full numerical model simulations would be required to test this idea. Such analysis would be highly worthwhile because σ is available from the Greenwich data from 1874 onwards and it could be used to define the emerged variation since that date, if the association suggested do apply.

The potential relevance of this work to TSI variability lies in the fact that it is the distribution of flux tubes, and the total flux which they contain, which has modulated the surface emissivity and TSI over recent solar cycles. For example, *Lean et al.* [2002] argue that because they can model a long-term drift in open flux (panel b of figure 3.20), of the same magnitude as derived from geomagnetic data by *Lockwood et al.* [1999a], using only flux emerged in large tubes (spots), and it can be seen from panel c of figure 3.20, the surface flux returns to almost constant, near zero values at each minimum. From this they argue the TSI returns to a near constant value at each sunspot minimum with little long-term drift. However this argument does not allow for any emergence of smaller flux tubes in an extended solar cycle in which varying degrees of overlap would allow a long-term drift in the solar minimum surface flux and in the TSI. *Solanki et al.* [2002] have allowed for such an extended solar cycle in the emergence of small flux tubes and found a good match with the *Lockwood et al.* [1999a] open flux variations. (NB this reduces the time constant T required in the open flux continuity equation to nearer 2 years). The drift in sunspot spread shown in this chapter suggests that an extended solar cycle in small flux tube emergence, between the sunspot groups, may well be present. By having variable overlap this would give a long-term drift in surface flux and thus, potentially TSI. This point will be discussed in Chapter 6.

Chapter 4: Theory of solar luminosity, irradiance and radius variations.

4. 1. Introduction.

The Sun's luminosity and radius is known to vary over the ten year timescale of the solar cycle because of solar magnetic fields and on much longer timescales of $10^6 - 10^8$ years, as a consequence of stellar evolution, where the burning of hydrogen into helium in the solar core causes secular variations [*Schröder et al.,* 2001]. Our knowledge of the solar cycle variations comes from observations by high resolution instruments, many in space, over the past 25 years. The understanding of the secular change, on the other hand, comes from surveys of astronomical data on stars. For climate changes, variations on timescales of $10 - 10^3$ years are of particular interest. Because these timescales are considerably shorter than the time constants for energy transfer from the Sun's core to the surface and for warming of cooling the convection zone, the changes are likely to be magnetic in origin, as they are over the solar cycle. The variations in luminosity and radius associated with magnetic fields may be due to processes that occur in the solar convection zone or may be associated with modulation of the surface emissivity by magnetic fields in the solar surface. Large-scale photospheric flux tubes (above a threshold radius of about 250 km) will cause sunspots to appear on the solar surface, leading to darkening in the affected region, and a loss of luminosity and irradiance. Smaller scale photospheric magnetic flux tubes (below about 250 km radius) cause faculae to appear on the solar surface causing localised brightening, and an increase in solar luminosity and irradiance. In this chapter I resume my introduction with a review *Spruit's* theory of the effects of both these surface magnetic fields and those, which occur deeper in the convection zone, on solar luminosity and radius.

Surface effects: These are magnetic fields, which give a visible contribution to the solar output by modulating the emissivity of the surface. An example of this is the reduction of upward heat flux just under the surface in sunspots, which reduces their temperature from the normal surface value near 6000K to about 4000K, causing

them to radiate less than the surrounding photosphere. Faculae are similar magnetic flux tubes, the main difference being their smaller radius (less than about 250 km), and that they are ~100K hotter than the surrounding photosphere and are thus brighter, especially near the solar limb.

Subsurface effects: These can be split into two separate phenomena, which affect the convection zone [*Spruit* 1991]:

1. Shadows: These are changes in heat transport coefficient. Magnetic fields in the convection zone can interfere with convection, causing a reduction in the efficiency of heat transport towards the surface. this is called the "alpha effect".

2. Sources and sinks: The creation of a magnetic field involves the conversion of energy of motion into magnetic energy. Since the motions in the solar envelope are thermally driven, this ultimately means conversion of thermal energy into magnetic energy. Building up a magnetic field will produce a thermal sink somewhere. When the field decays the opposite will happen, and magnetic energy will be converted into heat energy. This is called the "beta effect".

4. 2. Timescales.

Thermal disturbances in the convection zone evolve on two different timescales [*Spruit,* 1982b, 1991]. The longest of these timescales is the thermal timescale (T_t) of the convection zone as a whole (which is also called the Kelvin-Helmholtz time scale). This is the timescale for warming or cooling the entire convection zone and is of the order 10^5 years. So for time scales of human interest, the convection zone has an extremely large thermal inertia. Even if the central heat source of the core of the sun were to be switched off completely, the internal thermal structure and surface luminosity would start to change only on this extremely long timescale.

The thermal time scale can be defined as a function of depth z, by considering the time taken for a heat flux F to take away the internal energy stored in a scale height H, HU:

$$T_t(t) = \frac{H(z)U(z)}{F(z)} \tag{4.1}$$

where H is the pressure scale height, U the thermal energy per unit volume at a depth z, and F the energy flux at a depth z. Because the heat flow into the surface must equal the heat flux radiated by the surface in steady state, the luminosity, L, equals the heat flux at the surface, F(z = 0), multiplied by the surface area, $4\pi R_s^2$. $T_t(z)$ is the time scale is on which the convective heat flux profile, and the observed surface luminosity, would start changing when the heat flux in the convection zone was interrupted at a depth z. The thermal time scale is a strong function of depth, due to the rapidly increasing temperature and density, and some rough values for T_t are, 10^5 yr at the base of the convection zone (z = $2x10^5$ km), 10 years at a depth of 16,000 km, and 10 hours at 2000km. So it can be seen that speed of the thermal response of the Sun depends critically upon the location of the magnetic disturbance.

The second time scale involved in thermal changes is that of the diffusive time scale T_d. This is the time scale on which differences in entropy between different parts of the convection zones are equalled out.

$$T_d = \frac{d^2}{K_t} \tag{4.2}$$

Where d is the size of the volume considered, and K_t is the turbulent diffusivity (from the mixing length theory it can be estimated that K_t is of order $10^9 m^2 s^{-1}$ at all z). For the base of the convection zone, z = $2x10^5$ km, T_d is about 1 yr, for z = 2000 km about 1 hr. At the surface (in the surface boundary layer) T_t and T_d are of a similar magnitude, but as we move into deeper layers, the thermal time scale increases rapidly and at the bottom of the convection zone, the thermal time scale is longer than the diffusive timescale by up to 10^5 years. Since the thermal time scale is so much larger than the diffusive timescale (even at z = 2000km, $T_d \approx T_t/10$), then it

can be considered that the changes even below the surface do not occur in thermal equilibrium. This means that heat flux blocked by spots can be stored in the convection zone.

4. 3. The heat flow equation.

The thermal adjustment of the convection zone can be described by the energy equation from the first law of thermodynamics [*Spruit*, 1982a,b, 1991]:

$$\rho T \frac{dS}{dt} = -\nabla . F + G \tag{4.3}$$

where ρ is the density, T the temperature, S the entropy per unit mass and F is the energy flux (convection plus radiation). G includes sources and sinks of heat, but gravitational energy is not included in G, so the thermal effects can be calculated without the gravitational effects being taken into consideration.

In the mixing length-approximation F can be written as:

$$F = -K_t \rho T \nabla S \tag{4.4}$$

giving F as a function of the local entropy gradient ∇ S.

Using (4.4),we can write (4.3) as:

$$\rho T \frac{dS}{dt} = K_t \nabla . (\rho T \nabla S) + G \tag{4.5}$$

where K_t is assumed to be constant, since it is approximately independent of depth.

A quasi-hydrostatic approximation is introduced to describe the change in pressure due to the local acceleration due to gravity:

$$\frac{dP}{dz} = g\rho \tag{4. 6}$$

where P is the gas pressure, and z is depth below the surface [*Spruit*, 1991].

If we now assume that the convection zone is thin enough, then g will be constant, which is a good analogy for the Sun, equation (4.6) then has a solution of form

$$P = P_o e^{\frac{z}{H}} = P_o e^{\mu} \tag{4.7}$$

where μ is a convenient depth parameter.

$$\mu = \ln\left(\frac{P}{P_o}\right) \tag{4.8}$$

where P_o is the reference gas pressure at the surface (z = 0, μ = 0). Equation (4.5) can now be rewritten as:

$$H^2 \frac{dS}{dt} = K_t \frac{\partial^2 S}{\partial \mu^2} + K_t(1-\nabla)\frac{\partial S}{\partial \mu} + \frac{H^2}{\rho T} G \tag{4.9}$$

where H = dz/dμ is the pressure scale height, and $\nabla = \partial \ln T / \partial \mu$ is the logarithmic temperature gradient [*Spruit*, 1982b, 1991]

If we go back to equation (4.5) and we neglect sources and sinks (G) and reduce to vertical variations, (so the operator ∇ becomes $\partial / \partial z$), we get:

$$\frac{dS}{dt} = K_t \frac{\partial^2 S}{\partial z^2} + \frac{K_t}{H}\frac{dS}{dz} \tag{4.10}$$

where $H=[\partial \ln(\rho T)/\partial z]^{-1}$ is the pressure scale height. If the first term on the right hand side dominates then:

$$\frac{dS}{dt} = K_t \frac{\partial^2 S}{\partial z^2} \tag{4.11}$$

which is a standard diffusion equation, and it's solution will evolve on the diffusive time scale which, from the form of (4.11) is (z^2/K_t), as quoted in equation (4.2).

If the second term in (4.10) dominates then:

$$\frac{dS}{dt} = \frac{K_t}{H}\frac{\partial S}{\partial z} \tag{4.12}$$

and we use the expression for $S = c_v(\ln P - \gamma \ln \rho)$ (c_v and c_p are the specific heats at constant pressure and volume, respectively, the ratio of which is γ) and using equation (4.4), equation (4.12) becomes:

$$\frac{d}{dt}(\ln P - \gamma \ln \rho) = -\frac{F}{\rho T c_v H} = -\frac{F}{UH} \tag{4.13}$$

where the internal energy per unit volume, U, is $\rho T c_v$ [*Spruit*, 1991]. This gives a time constant of HU/F, as given in equation (4.1).

4. 4. Polytropic model.

As discussed above, magnetic fields can do one of two things: either introduce a new energy source/sink by magnetic flux being destroyed or created in the convection zone (beta perturbation), or can change the energy transport coefficient (alpha perturbation). To understand these two separate effects we need to model the convection zone. In order to solve the heat transport equations, *Spruit* [1976, 1982a,b, 1991] used a pseudo-polytropic model of the convection zone, which is a linear variation of temperature with depth. The depth dependence can be described for the following quantities as:

$$\zeta = 1 + z/((n+1)H_0) \tag{4.14a}$$

$$P = P_0 \zeta^{n+1} \tag{4.14b}$$

$$\rho = P_0/(gH_0)\zeta^n \tag{4.14c}$$

$$H = H_o \zeta \tag{4.14d}$$

where ζ is the depth variable and n is an index. A good fit for the Sun's convection zone is n = 2, H_o = 1.5 x 10^7cm, P_o = 4 x 10^4 Pa. Because the surface gravity is g_o=274 ms^{-2}, this gives a surface density of $P_o/(g_oH_o)$ of 10^{-3}kg m^{-3}. The logarithmic temperature gradient is fixed by the value of n:

$$\nabla = \frac{\partial \ln T}{\partial \mu} = \frac{\partial \ln T}{\partial \ln P} = \frac{1}{n+1} \tag{4.15}$$

Spruit [1982b] shows that the solution of the equation (4.9) using the polytropic model is that a small fractional temperature perturbation (T'/T) gives a heat flux perturbation F' given by:

$$\frac{F'}{F_{\square}} = \frac{-T}{T_0} + \frac{\zeta^{n+1}}{\delta_0 (n+1)} \frac{\partial \left(\frac{T'}{T} \right)}{\partial \zeta} \tag{4.16}$$

where:

$$\delta_0 = \frac{F_0 H_0}{T_0 K_t \rho c_p} \tag{4.17}$$

Equation (4.16) applies at all depths of the convection zone up to the base of a surface layer. *Spruit* [1982b] showed the solution cannot also apply in a thin "superadiabatic" surface layer and so the convection zone model given by equation (4.16) must be used in conjunction with a thin emitting surface layer model.

4. 5. The Surface boundary layer

At the surface the temperature is T_s and the heat flux is F_s, which by conservation of energy must be continuous with the heatflux in the underlying convection zone [*Spruit*, 1982a, b]. If we assume a blackbody radiation $F_s = \sigma T_s^4$ (the Stefan-Boltzman law) and differentiate:

$$\frac{dF_s}{dT_s} = 4\sigma T_s^3 = 4\frac{F_s}{T_s} \tag{4.18}$$

using the notation $F_s' = dF_s$ and $T_s'=dT_s$:

$$\left(\frac{F_s'}{F_s}\right) = 4\left(\frac{T_s'}{T_s}\right) \tag{4.19}$$

The surface temperature T_s will, in general depend on the solar radius, R_s, the heat flux F_s at the surface, and S_o, the entropy at the base of the surface layer. For small perturbations we can write:

$$T_s' = \left.\frac{\partial T_s}{\partial R}\right|_{F,S_o} R' + \left.\frac{\partial T_s}{\partial F}\right|_{R,S_o} F' + \left.\frac{\partial T_s}{\partial S_o}\right|_{R,F} S_0' \tag{4.20}$$

to calculate the first term, we study the dependence of the boundary layer on surface gravity [*Spruit*, 1991]. The mixing-length model gives a dependence $\partial \ln T / \partial \ln g \approx 0.3$. This number is determined mostly by the dependence of opacity at the surface on temperature and density. Since $g \sim R^{-2}$ we find:

$$\left.\frac{\partial T_s}{\partial R}\right|_{F,S_o} \approx \frac{0.6T_s}{R_s} \tag{4.21}$$

this term can be seen to depend upon the solar radius. The last two terms can be calculated from the polytropic model solution (equation 4.16) [*Spruit*, 1991], at a surface where ζ=1:

$$\frac{\partial T_s}{\partial F_s} = \frac{T_s}{F_s}\left[\exp(-\frac{2}{3}\delta_0) - 1\right] = \frac{T_s'}{F_s'} \tag{4.22}$$

for constant pressure $S_O' = c_p(T_o'/T_o)$, thus:

$$\frac{\partial T_s}{\partial S_0} S_0' = T_s\left(\frac{T'}{T_0}\right) \tag{4.23}$$

which is the temperature perturbation at the top of the envelope. From (4.19) we get:

$$\frac{F_s'}{F_s} = -0.6\eta\frac{R_s'}{R_s} + \eta\left(\frac{T_0'}{T_0}\right) \tag{4.24}$$

where:

$$\eta = \left[\frac{5}{4} - \exp\left(-\frac{2}{3}\delta_0\right)\right]^{-1} \tag{4.25}$$

Because n = 2, $\delta_o \approx 0.25$ and $\eta \approx 1.8$. Equation (4.24) gives the surface flux change if changes in surface radius and temperature are known. Since luminosity, $L = 4\pi R_s^2 F$ then:

$$\frac{dL}{dt} = 8\pi R_s\frac{\partial R_s}{\partial t}F_s + 4\pi R_s^2\frac{\partial F_s}{\partial t} \tag{4.26}$$

$$L' = 4\pi R_s^2 F_s\left(\frac{2}{R_s}R_s' + \frac{F_s'}{F_s}\right) = L\left(\frac{2R_s'}{R_s} + \frac{F_s'}{F_s}\right) \tag{4.27}$$

substituting (4.24) yields:

$$\frac{L'}{L} = (2 - 0.6\eta)\left(\frac{R_s'}{R_s}\right) + \eta\left(\frac{T_0'}{T_0}\right) \tag{4.28}$$

so from this we can look at the effects on the luminosity of radius changes and temperature changes at the base of the surface layer [*Spruit*, 1991]. I will return to this in section 4.7 and look at the effect of blocked upward heat flux on the solar radius and, via equation (4.28) its contribution to luminosity change.

4. 6. Effect of blocked heat flux.

Not all heat that is blocked is stored in the convection zone, and a fraction α will re-appear at the surface, depending upon the depth of the blockage. Convective flows and the transport of heat by turbulent diffusion may result in more of the blocked heat emerging in bright rings around spots than would be expected from *Spruits* equations. Bright rings around spots have recently been studied using precision photometry by *Rast et al.* [1999]. (See figure 4.1).

On a timescale t which exceeds the thermal timescale for the depth d of the heat block in the spots, $t > T_t(d) > T_d(d)$, then both the heat flux profile and the temperature profile above the base of the spots ($z<d$) will have adjusted to the presence of the spot. Just below the sunspots ($z=d$) the average upward heat flux is F_L and in the surface in a layer effected by sunspots ($z<d$) the average heat flux is F_s. In the absence of spots, $F_L = F_{SP}$ everywhere and the luminosity, $L = 4\pi R_s^2 F_L$. In figure 4.2 the upper layer is divided into three classes when sunspots are present: (1) undisturbed sun, covering a fraction of the f_o of the surface and through which the heat flux is F_{SP}; (2) Dark spots, taken to be here an average of umbra and penumbra and covering a fraction f_S of the surface and through which the heat flux is F_{SS}; (3) Bright rings around spots, covering a fraction f_R of the surface and through which the heat flux is F_{SR}. The average flux through the upper layer is then:

$$F_S = f_0 F_{SP} + f_S F_{SS} + f_R F_{SR} \qquad (4.29)$$

where:

$$f_o + f_S + f_R = 1 \tag{4.30}$$

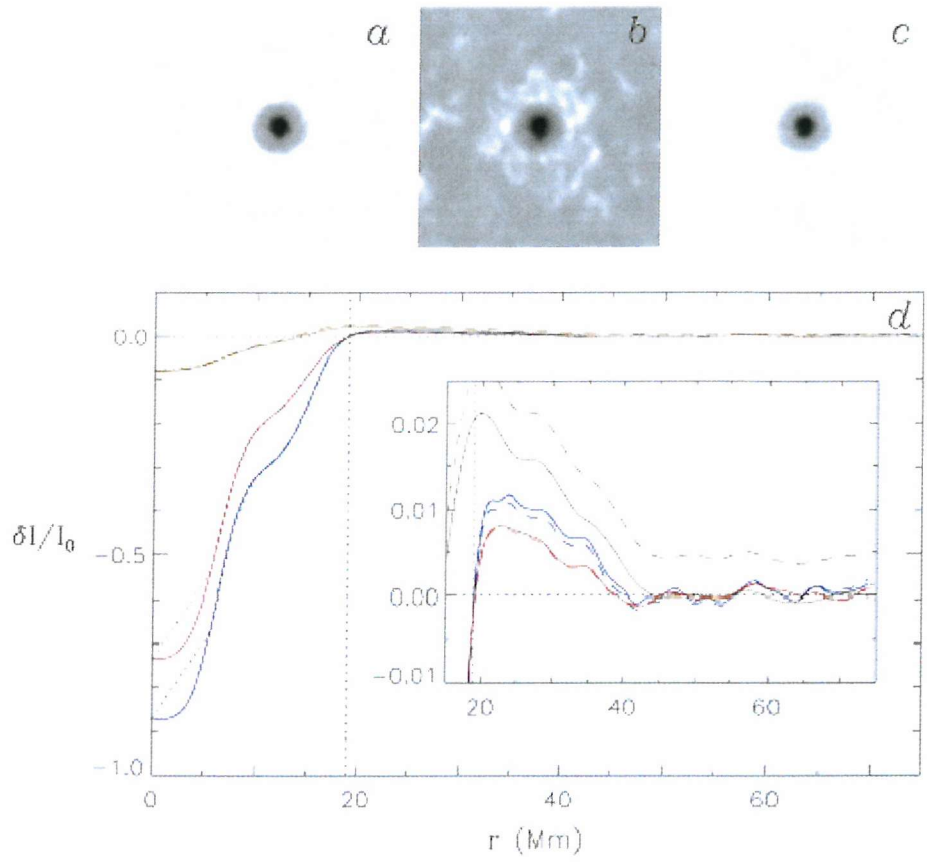


Figure 4.1. Observations of the intensity in and around a sunspot by *Rast et al.* [1999]. Images (a), (b) and (c) show a 150 Mm × 150 Mm area around sunspot NOAA 8263 as observed through, respectively, the blue, Ca II K, and red filters of the PSPT (Precision Solar Photometric Telescope) on 6 July 6, 1998. (d) Azimuthal averages of the residual intensity (given as $\delta I/I_o$, where $\delta I = I(r) - I_o$, and $I(r)$ and I_o are the intensities at r and of the undisturbed photosphere) as a function of distance r from the spot centre for all three wavelengths (blue, red, and brown curves, corresponding to blue continuum, red continuum, and Ca II K intensities, respectively). Inner box shows close up of figure (d) starting at r = 20 Mm.

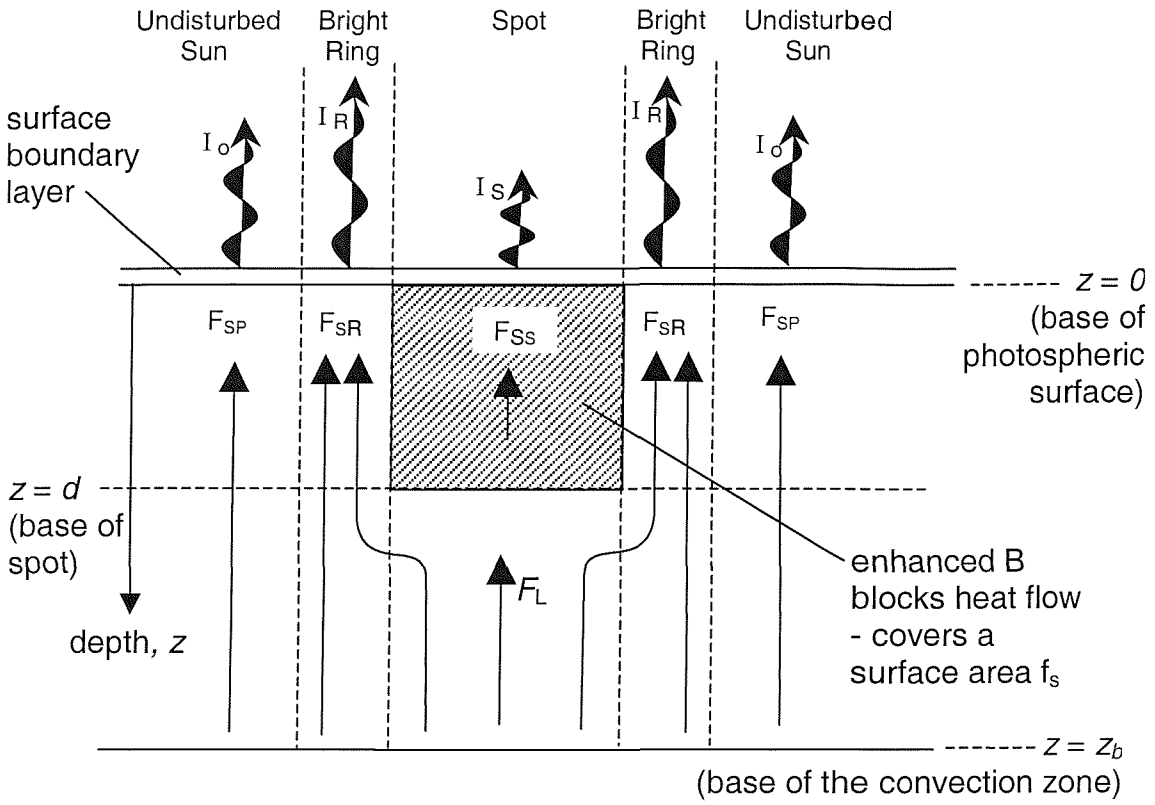


Figure 4.2. Schematic of heat blocking by an enhanced field region below a sunspot in the surface layer.

α is defined as the fraction of the blocked heat flux that still reaches the surface. The blocked heat flux is $f_S(F_{SP}-F_{SS})$ and the heat flux returned to the bright ring is $f_R(F_{SR}-F_{SP})$. Thus:

$$\alpha = \frac{f_R\left(F_{SR}-F_{SP}\right)}{f_S\left(F_{SP}-F_{SS}\right)} \tag{4.31}$$

From (4.31, 4.30 and 4.29)

$$\frac{L'}{L} = f_S\left(1-\frac{F_{SS}}{F_{SP}}\right)(1-\alpha) \tag{4.32}$$

From figure 4.1 it can be found that $(f_R/f_S) \approx (\pi(2r_s)^2-\pi r_s^2)/\pi r_s^2 = 3$. (where r_s is the spot radius), and that the average (by area, rather than radius) over the ring is $F_{SR} \approx 1.03F_{SP}$. From 4.31 this gives α of 0.09, roughly 10% of the blocked flux reaches

the surface. It should be noted that the fraction (1 – α) that does not reach the surface is stored in the deeper layers of the convection zone. This maybe a relatively large value for α as many spots show no such strong bright ring and α is lower in these cases. As will discussed in section 4.8.2, the α may depend on the age of the spot.

It is useful to compare this value for α with the solution of the polytropic model. *Spruit* [1976, 1982b] derives an expression:

$$\alpha = 1 + \left[\left\{ \exp\left(\frac{-(n+1)}{n} \delta_0 \zeta_d^{-n} \right) - 1 \right\} \left\{ \frac{5}{4} \exp\left(\frac{n+1}{n} \delta_0 \left(1 - \zeta_d^{n} \right) \right) - 1 \right\}^{-1} \right]^{-1} \quad (4.33)$$

where, from the definition of ζ,

$$\zeta_d = 1 + \frac{d}{(H_0)(n+1)} \quad (4.34)$$

If we consider spots deep enough so that d » (n+1) H_o, then d / [(n + 1) H_o], so that:

$$\alpha \approx \zeta_d^{-n} \left(\frac{n+1}{n} \right) \delta_0 \left\{ \frac{5}{4} \exp\left(\frac{(n+1)}{n} \delta_0 \right) - 1 \right\}^{-1} \quad (4.35)$$

If we take a value of n = 2 for the convection envelope and δ_o = 0.25, then:

$$\alpha \approx 0.5 \zeta_d^{-2} = 0.5 \left(1 + \frac{d}{3H_o} \right)^{-2} \quad (4.36)$$

If a value of d = 3500km and $H_o = 1.5 \times 10^5$ m, then equation (4.36) yields α ≤ 0.6%, which is smaller than the 9% inferred from bright rings and it requires a depth of just d = 612 km to yield α = 9%. Thus either spots are much shallower than we have inferred from helioseismology data, or the combination of turbulent and

diffusion and convection flows may be more efficient in bringing heat flux to the surface than the above theory predicts.

4. 7. Effect of Radius Changes.

Due to the appearance of spots at the surface of the Sun, the temperature outside the spots is slightly increased. Hydrostatic equilibrium then requires that the stellar radius measured outside the spots is slightly greater. Inside the spots the temperature is lower, and so the local stellar radius is reduced (the Wilson depression), so we have to distinguish between the radius change inside and outside the spots [*Spruit*, 1982b, 1991]. We can calculate the radius changes outside the spots from hydrostatic equilibrium:

$$R_s' = \int_0^D \frac{T'}{T} dz \tag{4.37}$$

for small T'/T, and for $\delta_o < 1$ we find:

$$R_S' = \frac{9}{2}\delta_0 H_0 (1-\alpha) f_S \frac{3dH_0}{(d+3H_0)^2} \tag{4.38}$$

For very deep and very shallow spots, R_S' is negligible, and R_S' has a maximum value when $d= 3H_o$, such that:

$$R_S' < \frac{9}{8}\delta_0 H_0 (1-\alpha) f_S \tag{4.39}$$

if we take $\alpha = 0.125$ (for the depth of $d = 3H_o$), and $f_S = 0.1$, we find that the radius change (R_s') outside the spot is 3.7×10^3 m, which is extremely small, when compared to the radius of the sun as a whole, and so will be unmeasurable.

If we take F_{SS} to be zero, 4.32 reduces to $L'/L = -(1-\alpha) f_S$, we can calculate the ratio of the relative radius change to relative luminosity changes by substituting into 4.39:

$$W = \frac{R_S'/R_S}{L'/L} \langle -\frac{9}{8}\delta_0 \frac{H_0}{R_s} \tag{4.40}$$

for the sun we find that $W = -6\times10^{-5}$, and so radius changes are negligibly small compared to the luminosity changes [*Spruit*, 1982b].

If we take the depth of spots to be 1000 km deep, then $\alpha = 0.05$, then the surface temperature inside spots = 4100K (T_{ss}), and the surface temperature outside the spots = 5770K (T_{sp}), then the ratio of heat flux inside and outside the spots is:

$$\frac{F_{ss}}{F_{sp}} = \frac{\sigma T_{ss}^{\ 4}}{\sigma T_{sp}^{\ 4}} = \left(\frac{4100}{5770}\right)^4 = 0.25 \tag{4.41}$$

so spots radiate at one quarter of the rest of the undisturbed photosphere. So the heat flux blocked = 0.75 F_{sp} and of that, up to 10% still makes it to surface, 0.075 F_{sp}. Therefore the maximum heat flux making it to the surface is (0.25 +0.0075) F_{sp} = 0.325 F_{sp}. If we look at the luminosity of the Sun when clear of spots and covered with a fraction f_S of spots:

$$L_{spots} = 4\pi R_s^{\ 2} f_S F_s = 4\pi R_s^{\ 2} f_S 0.325 F_{sp} \tag{4.42}$$

$$L_{\text{no spots}} = 4\pi R_S^2 F_{SP} \tag{4.43}$$

which therefore gives a ratio of the luminosities of spotted and unspotted sun:

$$\frac{L_{spots}}{L_{\text{no spots}}} = 0.325 f_S \tag{4.44}$$

If we take f = 0.3%, which is a reasonable amount for the sun at solar maximum, then the ratio of the luminosities is 0.1 %, given $f_S \approx 0$ at sunspot minimum this predicts that the effect of sunspots is to make the sun 0.1% less luminous at sunspot maximum than at sunspot minimum. This is similar to the decreases in TSI due to sunspots that have been calculated by *Fröhlich et al.,* [1994]

4. 8. Temporal development of the effects of Magnetic Fields

As mentioned at the start of this chapter, magnetic fields below the surface of photosphere have two effects

1: Creation/destruction of magnetic fields at a depth z will cause a sink or source of energy. This is called the **β** effect.

2: By interfering with convection motions, a magnetic field can locally increase the entropy gradient needed for transporting energy, in effect influencing the energy transport coefficient. This is called the **α** effect.

4. 8. 1. The β Effect.

If we assume that magnetic flux is created or destroyed at a depth z_b, in a thin layer dz thick and on a timescale t, where $T_d < t < T_t$. then the rate of growth of magnetic energy is:

$$\frac{d}{dt}\left\{\left|\frac{B^2}{2\mu_0}\right| dz\right\} = G \qquad (4.45)$$

where G is the source term which first appeared in equation (4.3). If we solve the heat balance equation above and below z_b:

$$F_{(z\rangle z_b)} - F_{(z\langle z_b)} = \frac{d\left\{dz\dfrac{B^2}{2\mu_0}\right\}}{dt} \tag{4.46}$$

since the timescale are long, the solution will involve the thermal mode. Upon a sudden increase of B ($G > 0$, i.e. an energy sink), L initially rises due to expansion of R_s under the enhanced magnetic pressure (this effect is also seen at large t). At intermediate t, L is reduced because of the cooling at $z = z_b$ which spreads to the surface on the diffusive timescale timescale $T_d(z_b)$. The temperature perturbation T'/T is constant with depth (except in the superadiabatic surface layer) and so the temperature must be continuous across z_b. The solution for the heat flux can be seen in figure 4.3, and it can be seen that the heat flux which is destroyed in the sink, is almost entirely supplied from below the level z_b [*Spruit*, 1991]. A consequence of this is that only a very weak signal (if any) reaches the surface. A signal of any magnitude will only reach the surface if the sink varies fast enough (the diffusive timescale), so there is very little luminosity change associated with such sinks. The longest timescale is for effects at the base of the convection zone and is about one year.

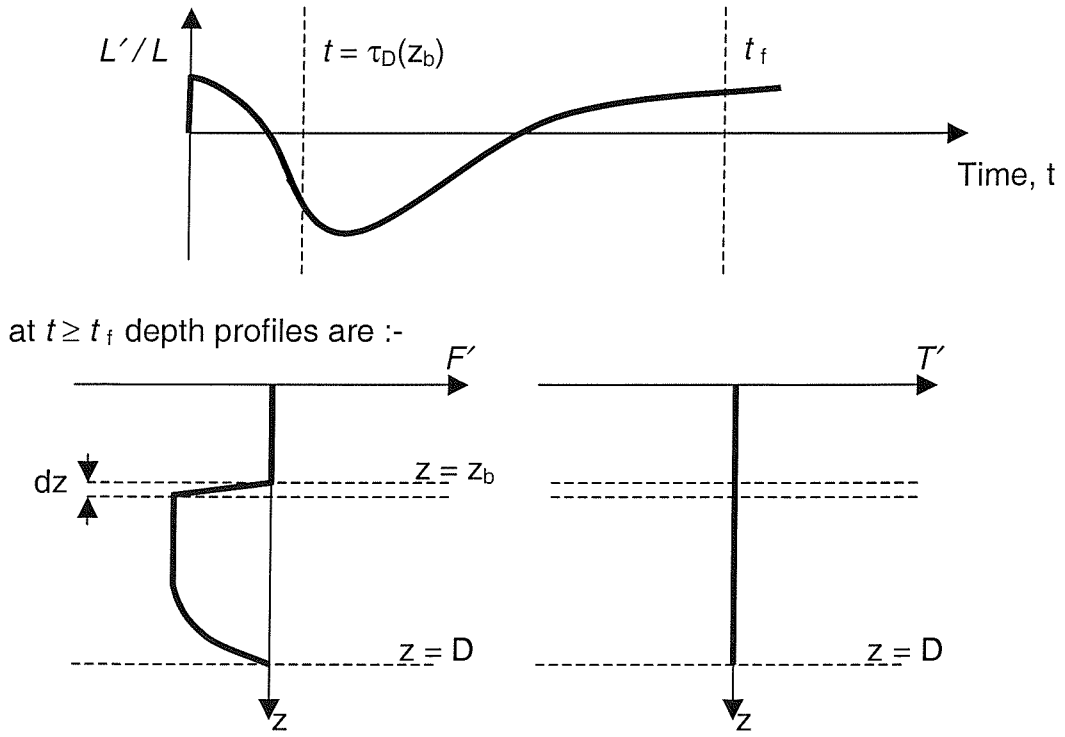


Figure 4.3. Predictions of the β effect, where the field B changes in a slab, dz thick at depth z_b. (Top) the time variation in the fractional luminosity perturbation L′/L. (Bottom left) the profile of the perturbation in heat flux, F′ at $t = t_f$. (Bottom right) the profile of the perturbation in temperature, T′ at $t = t_f$. [*Spruit*, 1991]

4. 8. 2. The α effect.

In the **α** effect model the magnetic energy density is compared with the kinetic energy density in convective turbulence. When these two energy densities are comparable, the magnetic field has the equipartition strength, **B**e defined by $1/\delta \approx \beta = 2\mu_o P/B_e^2$. The magnetic field is now strong enough to start reducing the degrees of freedom of the convective flow, so a larger entropy gradient is needed to transport the same energy flux [*Spruit*, 1991]. To include this effect we modify the mixing-length expression for the convective energy flux. Equation 4.4 can be re-written using $H = c_p(\nabla - \nabla_a) = c_p\delta$ and for radial stratification, $S_o \nabla S = \partial S / \partial z = c_p \delta / H$. This yields $F = K_t \rho c_p (T/H)\delta$. This is modified to:

$$F_{conv} = K_t \rho c_p \frac{T}{H}(\delta - q/B) \qquad (4.47)$$

where q is a factor of unity.

If we turn on the α effect in a layer d wide with its upper layer at a depth of $z = z_b$, then F_{z_b} will be reduced, so T above z_b will start to cool, and T below z_b will rise. This can be seen in a series of diagrams showing the effects of a reduction in energy transport (figure 4.5). The first diagram shows that for $T_d(z_b) < t < T_t(z_b)$, where adjustments have occurred on the diffusive timescale, but before the layers above z_b have returned to equilibrium. The temperature below z_b has not changed, but the temperature above z_b is slightly reduced due to the mismatch between the heat fluxes at $z = 0$ and $z = z_b$. For timescale $T_t(z_b) < t < T_t(D)$, the layers above z_b have returned to thermal equilibrium, while the layers below are still slowly heating up, the surface flux is reduced during this period, and the temperature of the layers above z_b are reduced. On very long timescales $t > T_t(D)$, the whole convection zone is back in equilibrium, and the temperature below the layer has increased.

The time constant for each of these periods can be calculated. If $z_b = 10^6$ m, then $T_d(z_b)$ is 10^3 seconds ~ 15 minutes. For $T_t(z_b) = UH/F = 3.34 \times 10^3$ seconds = 1 hour, and at the base of the convection zone $T_t(D) = 10^5$ years. If we take a spot at a depth of 1000 km then the time series is displayed in figure 4.4. The bright ring seen

around spots first appears at surface at T_t (z_b), and reaches full luminosity at T_d (D), when the disturbance has propagated through the entire convection zone and the envelope enters a quasi-static phase, where the entire convection zone is heating up on the thermal timescale.

If the spot is switched off, the heat in the deepest layers is released. If the spot is switched off during the quasi-static phase then, the luminosity is enhanced to L_o (1+ αf), and decays to the normal luminosity on the diffusive timescale T_d. If the spot is switched off after a time T_t, the luminosity is enhanced to L_o (1 + f), and returns to the normal luminosity on the thermal timescale.

This raises an extremely interesting question. The typical lifetime of spots on the surface of the Sun is $T_s = 8.6 \times 10^6$ s, which is around 100 days, so what happens to the heat that is stored in the deepest layers, when this blockage is removed? By *Spruit*'s calculations the blocked heat that is stored in the convection zone should start to heat the surface on the diffusive timescale T_d (z_b) of the blockage, reaching a maximum brightening after time T_t (z_b), then decaying to the normal luminosity of the surface on the diffusive timescale of the convection zone T_d (D). This means that after time T_t (z_b) the surface luminosity will be enhanced to L_o (1 + αf). This enhancement could last for as long as a year. In effect, this may mean that some of the heat flux is blocked by the appearance of sunspots will re-appear later on, the heat flux in effect being stored and then released in the convection zone. This additional surface luminosity that this may cause could be in the form of ephemeral flux, network faculae, or quiet sun emissions

This leads to a possibility that a fraction of the lost emission in darkening caused by spots is later returned as brightening at the surface causing an increase in luminosity after the spots have dissipated. If the unblocking of the heat flux were delayed by sufficient time, the delayed return of heat flux to the surface could in part explain the increase in the quiet sun irradiance over the past 100 years. It has been postulated that solar minimum values of solar irradiance has steadily increased since 1900 (see chapter 6), but the reason for this is unknown. In general it is due to network faculae effects and/or quiet sun irradiance increases. It is possible that as the photometric sunspot index (see chapter 6) has increased over the past 100 years so

has the returned heat flux. As the PSI increases the amount of heat flux stored in the convection zone will also increase. *Spruit* [1982b] predicts that this stored heat flux will then be released after the spots have dissipated, causing brightening. The greater the amount of darkening, the greater the amount of delayed brightening. The one problem with invoking this as a cause of variation in the solar-minimum irradiance is that we would expect that the brightening caused by the unblocking would be greatest 1-2 years after the sunspot was seen. It is not clear that this is a sufficient delay to give a significant effect at sunspot minimum.

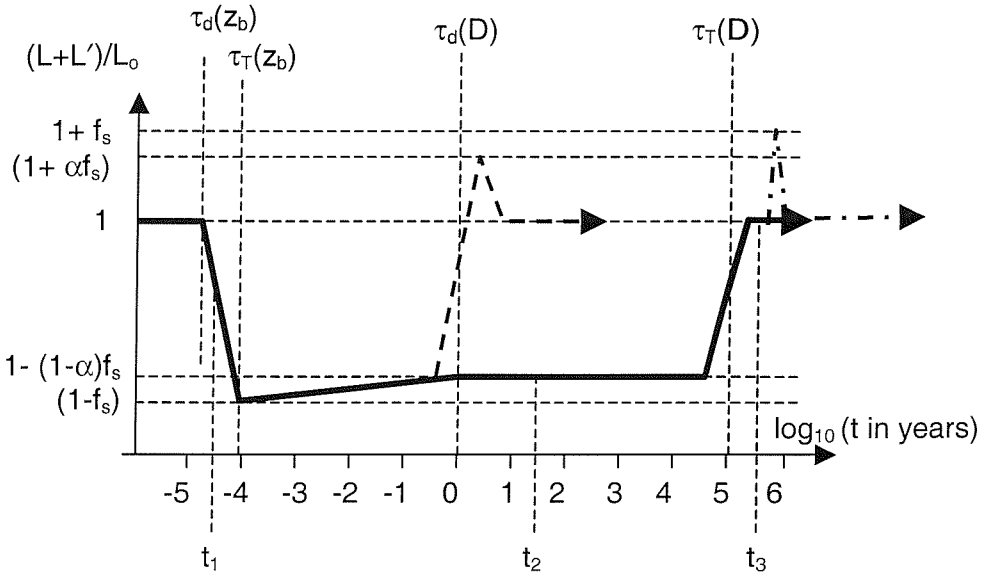
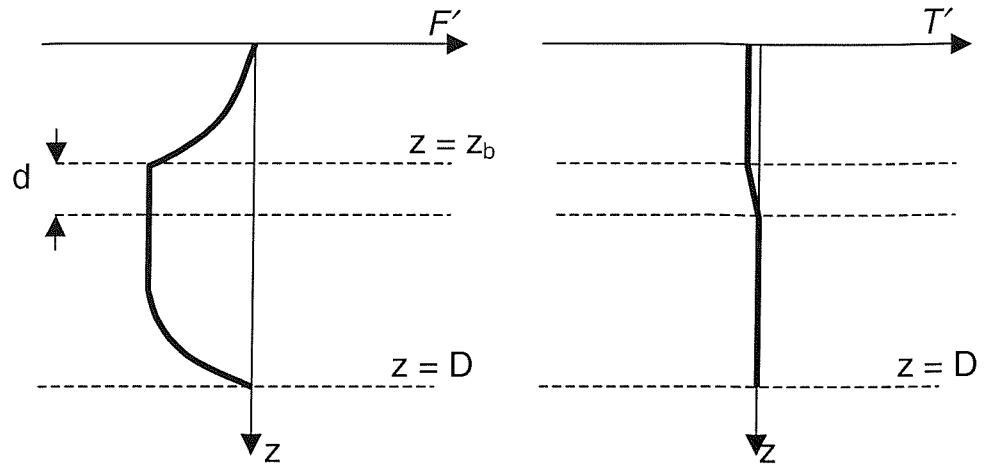
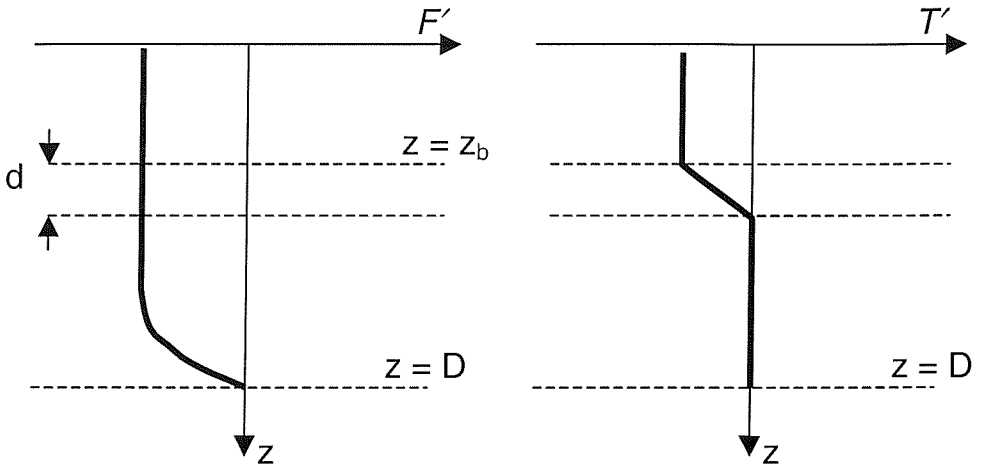


Figure 4.4. Predictions of the α effect: the time variation (on a log scale) in the fractional luminosity perturbation L'/L. The luminosities given assume that all flux is blocked. f_S is the fraction of the solar surface effected. The profiles at time t_1, t_2 and t_3 can be seen in figure 4.5.

at $t = t_1$ where $\tau_d(z_b) < t_1 < \tau_T(z_b)$, depth profiles are:-



at $t = t_2$, where $\tau_T(z_b) < t_2 < \tau_T(D)$ depth profiles are :-



at $t = t_3$, where $t_3 > \tau_T(D)$, depth profiles are :-

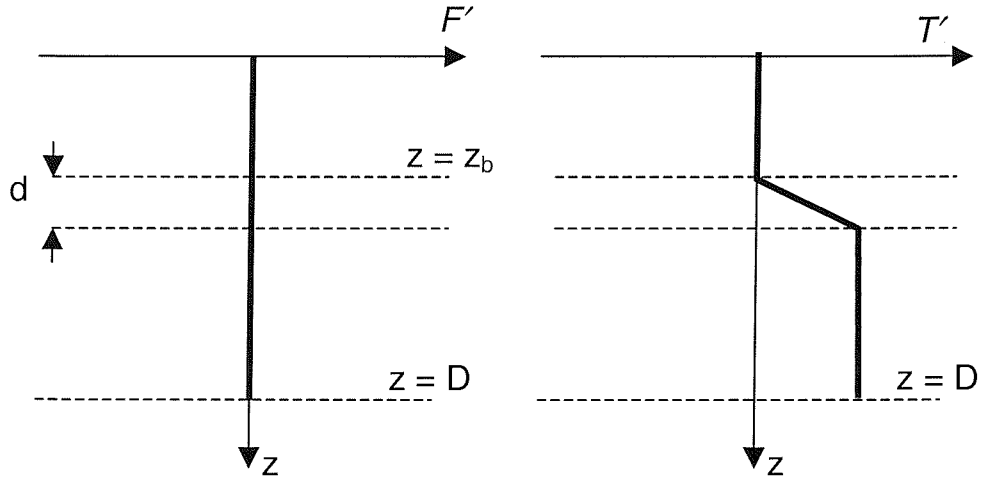


Figure 4.5. Predictions of the alpha effect. (Left) the profile of the perturbation in heat flux, F′ at various t. (Right) the profile of the perturbation in temperature, T′ at various t.

4. 9. Effects of surface magnetic fields.

In recent years, there have been several studies have been able to explain all observed variations in the total solar irradiance by summing the effect of all surface magnetic features [e.g. *Fligge et al.*, 1998; *Solanki et al.* 2002; *Krivova et al.* 2003]. These studies appear to allow, at most, small effects of alpha and beta phenomena of fields inside the convection zone, although variations in the intensity of the limb photosphere have, in the past, been explained in terms of such effects [*Libberecht and Kuhn*, 1984]. In this section, I look at the mechanisms of the surface effects in the light of the heat blocking theory outlined in the last section. I also look at how we can quantify the effect of the surface features.

4. 9. 1. Sunspot darkening: The Photometric Sunspot Index.

Sunspots occur due to a concentrated magnetic field that appears below the affected region (at a depth z_b) causing a blockage of the upward heat flux, as outlined in section 4.7. When the blockage first appears, convection zone above the blockage cools whereas below the blockage starts to heat up, albeit very slowly indeed. Effects become visible at the surface on the diffusive timescale ($T_d = z_b^2/K$), as the surface region above the blockage will become darker. At the same time a bright ring will start to appear around the dark region, whose contrast will depend upon the alpha value of the blockage (see section 4.5), and this bright ring will reach its full amplitude on the diffusive timescale of the convection zone. The decrease in emission from the affected region causes the well-known sunspot darkening, which is known to contribute to variations in solar luminosity over the solar cycle. This darkening is quantified by the *photometric sunspot index*, P_{SI} [*Willson et al.*, 1981, *Hudson et al.* 1982; *Fröhlich et al.*, 1994].

The depth at which the sunspot blocks the heat flux is not well known and is a keen point of interest. For the alpha value to be high enough to cause the bright rings seen around some sunspots, the depth z_b must be shallow, but we know that it must be larger than the observed Wilson depression, which is around 600 km deep. If we assume that the depth of the spot is around 1000 km, this will yield a value of alpha

of 0.6%, meaning that almost none of the blocked heat should be returned to the surface.

We quantify the darkening effect of sunspots using the photometric sunspot index (PSI), the derivation of which is given below. The intensity of a region of the Sun, I, is a function of the disk position parameter μ, given by

$$\mu = \cos\theta \tag{4.48}$$

where θ is the angle the region subtends with the Earth-Sun line at the centre of the Sun: μ = 1 at the centre of the disk and μ = 0 at the photospheric limb. The mean value of the intensity over the whole disk is:

$$\langle I \rangle_D = 2\int_0^1 I(\mu)\mu d\mu \tag{4.49}$$

The total intensity of the whole disk is:

$$\pi R_s^{\,2} \langle I \rangle_D = 2\pi R_s^{\,2} \int_0^1 I(\mu)\mu d\mu \tag{4.50}$$

The quiet-Sun intensity variation can be written as $I(\mu) = I_O L_D(\mu)$, where $L_D(\mu)$ is the limb darkening function and I_O is the intensity of the quiet Sun at the centre of the visible disc. The normal photosphere is darker at the limb because optical depth τ = 1, is reached at a greater height, where the photosphere is cooler. The irradiance is the power per unit area received at the Earth, a distance R_o = 1AU = 150×10^6 km, from the Sun. The quiet Sun irradiance is

$$Q_0 = \pi\left(\frac{R_s}{R_0}\right)^2 \langle I \rangle_D = 2\pi\left(\frac{R_s}{R_0}\right)^2 I_0 \int_0^1 L_D(\mu)\mu d\mu \tag{4.51}$$

The intensity of the umbra of a spot varies with the spot's μ-value, μ_S, as

$$I_u = I_{u0} g_u(\mu_s) \tag{4.52}$$

and for a spot penumbra it varies as

$$I_p = I_{p0} g_p(\mu_s) \tag{4.53}$$

If the area of the umbra is A_U and of a penumbra is A_P, then the total area of the spot is $A_S = A_U + A_P$. Note that these surface areas give areas on the visible disk of $\mu_S A_S$, $\mu_S A_U$ and $\mu_S A_P$. Then the spot changes the intensity by

$$\Delta I_s = \left(\frac{\mu_s A_u}{R_0^2}\right)\left[I_0 L_D(\mu_s) - I_{u0} g_u(\mu_s)\right] + \left(\frac{\mu_s A_p}{R_0^2}\right)\left[I_0 L_D(\mu_s) - I_{p0} g_p(\mu_s)\right] \tag{4.54}$$

To derive the PSI, it is assumed that all umbra have the same temperature, as do all pemnumbra and that all spots have the same ratio of their areas A_U/A_P. In addition, the limb darkening function is assumed to be the same for umbra, penumbra and the quiet sun, so $g_U(\mu) = g_P(\mu) = L_D(\mu)$, equation (4.54) then gives

$$\Delta I_s = \left(\frac{L_D(\mu_s)}{R_0^2}\right)\left[I_{s0} - I_0\right] \tag{4.55}$$

where:

$$I_{s0} = \left(\frac{A_p}{A_s}\right) I_{p0} + \left(\frac{A_u}{A_s}\right) I_{u0} \tag{4.56}$$

and ΔI_S is defined as positive for an intensity increase. From (4.51) and (4.55)

$$\frac{\Delta I_s}{Q_0} = \left(\frac{\mu_s A_s}{\pi R_0^{\,2}} \right) \left[\frac{L_D(\mu_s)}{2\int_0^1 L_D(\mu_s)\mu d\mu} \right] \left(\frac{I_{s0}}{I_{0-1}} \right) \tag{4.57}$$

We here define the spot contrast to be

$$c_s = \frac{I_{s0}}{I_0} - 1 = \frac{(I_{s0} - I_0)}{I_0} \tag{4.58}$$

(note that with this definition, positive/negative contrast corresponds to a brightening/darkening, respectively) and the filling factor (the fraction of the disk covered by the spot) is $\alpha_S = (\mu_S A_S / \pi R_o^{\,2})$. *Hudson et al* [1982] adopted the Eddington limb darkening profile

$$L_D(\mu) = \frac{(3\mu + 2)}{5} \tag{4.59}$$

this limb darkening profile is plotted as a thin dashed line, in figure (4.9), where it is compared to a more complex distribution by *Neckel and Labs* [1994],

Integration of (4.59), yields that the square term in brackets in (4.57) is equal to (3μ+2)/4. Thus from (4.57) and summing over all the spots present on the visible disk we get the photometric sunspot index (in Wm^{-2}):

$$\Delta I_s = -P_{SI} = Q_0 \sum_s \alpha_s c_s \frac{(3\mu_s + 2)}{4} = Q_0 \sum_s \left(\frac{A_s}{A_{sh}} \right) c_s \mu_s \frac{(3\mu_s + 2)}{2} \tag{4.60}$$

where A_{sh} is the area of a solar surface hemisphere and α_S is the disk filling factor. This is a definition of the photometric sunspot index, P_{SI}, which quantifies the effect of sunspots on the total solar irradiance. Note that sunspot contrasts are negative, by the definition used, and so the irradiance increase ΔI_S will be negative but P_{SI} is defined as positive. Because for monthly averages of P_{SI} longitudinal effects are

averaged out, the only influence of μ_S is through the latitudinal structure of sunspot occurrence. Thus influence is relatively small and so the variation of P_{SI} is dominated by that in the total sunspot area $\sum_S A_S$ on these timescales.

Estimates of the average temperatures are, $T_U = 4240$K, $T_P = 5680$K, and $T_{QS} = 6050$K, for umbrae, penumbrae and the quiet sun, respectively [*Allen*, 1973]. The fraction of the area of an average spot A_S is 0.18 for umbra and 0.82 penumbra ($A_U/A_S = 0.18$, $A_U/A_S = 0.82$).

Using the Stefan-Boltzman law for a blackbody radiator, so intensity I is proportional to T^4, the average contrast for a spot is

$$c_s = \left(\frac{A_u}{A_s}\right)\left\{\left(\frac{T_u}{T_{Qs}}\right)^4 - 1\right\} + \left(\frac{A_p}{A_s}\right)\left\{\left(\frac{T_p}{T_{Qs}}\right)^4 - 1\right\} = -0.32 \qquad (4.61)$$

in fact, c_S shows some dependency on spot size and position, and to generate PSI, *Fröhlich et al.* (1994) employ:

$$c_s = -0.2231 - 0.0244 \log_{10}\left(\mu_s A_s\right) \qquad (4.62)$$

We can get a rough estimate of the peak PSI by adopting the simple spot contrast given by (4.61). If spots are spread evenly over the surface, we can find the disc average of μ_S $(3\mu_S + 2)/2$ in equation (4.60) is:

$$2\int_0^1 \mu \frac{(3\mu = 2)}{2} \mu d\mu = 1.4167 \qquad (4.63)$$

monthly values of the total spot area $\sum_S A_S$ peak at sunspot maximum at about $(3\times10^{-3})A_{SH}$ with typical sunspot maximum values of about $(1.5\times10^{-3})A_{SH}$. From equation (4.58), the sunspot darkening $\Delta I_S / Q_o = 0.07\%$, which for $Q_o = 1365$ Wm^{-2}. This yields $\Delta I_S \approx 1$ Wm^{-2}.

In this thesis, we make use of the PSI data made available by the World Radiation Centre, PMOD, Davos on their ftp site (http://www.pmodwrc.ch). There are a number of second-order corrections to the simple formulation of PSI given by equations (4.60) and (4.62) which are included in these PSI data, as detailed by *Fröhlich et al.* [1994]. These corrections allow the PSI to accurately reproduce the observed effect on total solar irradiance of sunspot groups and individual sunspots as they rotate across the solar disk.

4. 9. 2. Facular brightening.

If the magnetic flux tube is smaller in diameter than a sunspot then it can emit more radiation than the surrounding photosphere, this is called a faculum ("torch"). There are two main theories of faculae: the hot wall model [*Spruit*, 1991; *Deinzer et al.*, 1984a; b; *Knölker et al.*, 1988; *Steiner et al.*, 1996] the bright cloud model [*Schatten et al.*, 1986]. In the hot wall model, faculae are very similar to sunspots, except that the radius of the flux tubes is smaller, allowing radiation from the tube walls to maintain the temperature: the increased optical depth inside the tube allows radiation from lower, hotter layers to escape, giving enhanced emission. The hot cloud model is dyanamical in that it considers the effect of upflows, which carry heat blocked in sunspots to the surface. A major difference between these models is the height of the surface in the faculae, compared to the surrounding photosphere – the hot wall predicting that the surface is depressed whereas the hot cloud model predicts that it is raised (so the latter is often referred to as the "hillock" model).

The Hot Wall Model

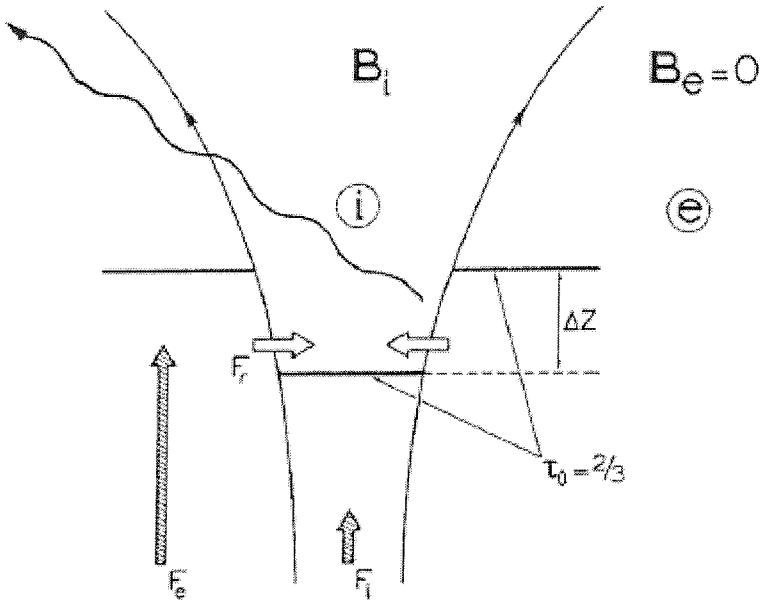


Figure 4.6. The hot wall model of faculae. The enhanced magnetic pressure in the flux tube means it has reduced particle pressure, but radiation from the walls (heat flux F_r) maintains the temperature even though the upward heat flux F_i is inhibited and reduced compared to the value F_e is outside the faculum. As a result, the constant optical depth (the $\tau_O = 2/3$ contour is shown here) is depressed by Δz. (From *Schrijver and Zwaan*, 2000).

Due to the pressure exerted by the magnetic field, the gas pressure inside the tube will be smaller than the surrounding photosphere at the same depth. This low pressure inside faculae will cause the opacity to be less than the surrounding area, so the optical depth unity will occur at a greater depth than for the surrounding photosphere. If we take a flux tube of $B_f = 1000$G, which is in pressure equilibrium with its surroundings, then:

$$\left(\frac{B_f^{\,2}}{2\mu_0}\right) + N_f k T_f = N_{Qs} k T_{Qs} \tag{4.64}$$

where T_f is the gas temperature in the faculae and T_{QS} is the temperature in the photospheric surface. If the tube is thin enough, then the horizontal exchange of

radiation ensures that $T_{QS} = T_f$ so that the concentration difference between the quiet photosphere and the faculum tube is:

$$\left(N_{Qs} - N_f\right) = \frac{B_f^{\,2}}{\left(2\mu_0 k T_{Qs}\right)} = 5\times10^2 m^{-3} \tag{4.65}$$

The $\tau = 2/3$ level will occur at a greater depth inside the faculum and thus the temperature will be higher and so more radiation is emitted. This will apply for all faculae which are smaller than the mean free path of photons at $\tau = 2/3$, which will be around 50 km, for tubes that are larger than this, the exchange of heat with the walls becomes less effective. Faculae of greater radius begin to behave in a similar fashion to spots at the disk centre where the centre of the tube will appear cooler than it's surroundings so the tube will not appear bright at the disc centre. Such tubes are often called micropores. The model requires the blocked heat flux to cause the tube floor to be cooled so that the flux tube appears dark when viewed from above. However, this tube would still contribute enhanced emission near the solar limb where the bright walls become visible.

The Hot Cloud Model

In this model, heat blocked by sunspots is conducted to the surface by magnetic flux tubes. At lower altitudes the upflow is mainly carried by upflowing protons which rise into the neutral hydrogen layer (which normally extends down to about 2000km below the surface). The recombination of the ionised hydrogen is exothermic, releasing additional energy and the gas is lifted by buoyancy. This forms a small bump or hillock, which is most visible on the limb of the sun.

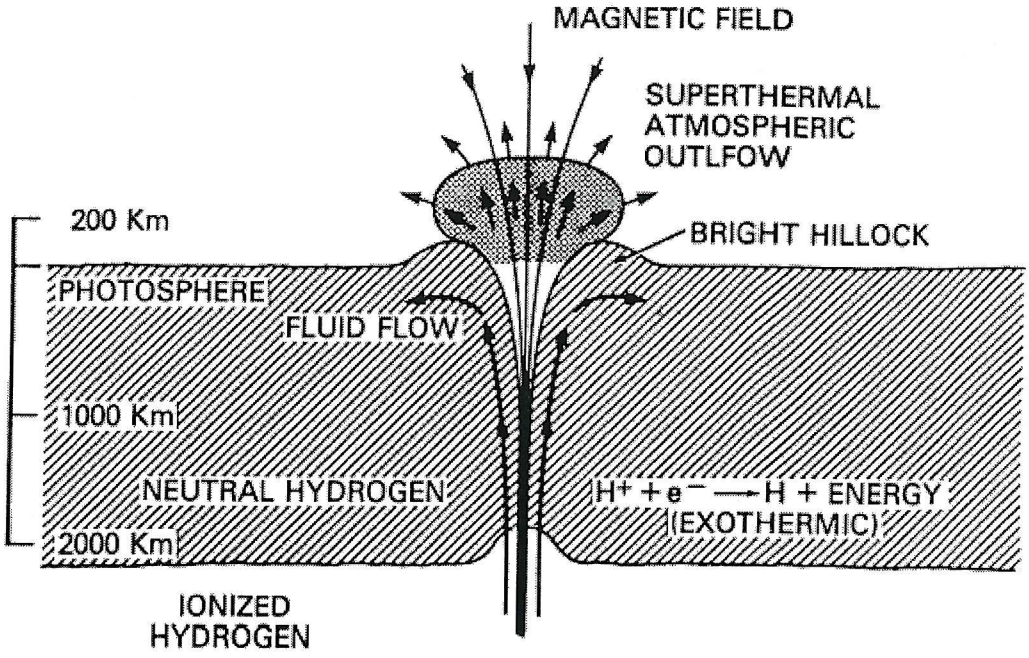


Figure 4.7. The hot cloud model of faculae. Upflows of hydrogen ions are driven up into the neutral hydrogen layer. These ions recombine exothermically, releasing more energy and driving the flows up and apart and so a bright hillock appears. [*Schatten et al.*, 1986].

4. 9. 3. Comparison of models.

Much evidence for the hot wall theory comes from the variation of contrast with the position parameter, μ. The hillock model predicts that faculae will be bright right out to the solar limb ($\mu = 0$), whereas in the hot wall theory predicts that the Wilson depression will cause faculae to vanish close to the limb. Much evidence of contrast, as a function of μ, has been interpreted as favouring the hot wall model. [*Topka et al.*, 1997; *Sánchez-Cuberes et al.*, 2003] and this model has gained widespread acceptance. However, there are problems, for example satisfactory explanation of the cool floor (and this lower contrasts at μ near unity) requires careful tuning of the model. In addition, recent very high-resolution observations by *Berger et al.* [2003] appear to show hillocks rather than dips. These observations were made using in 488 nm light using the Swedish 1-meter Solar Telescope on the island of La Palma, Spain.

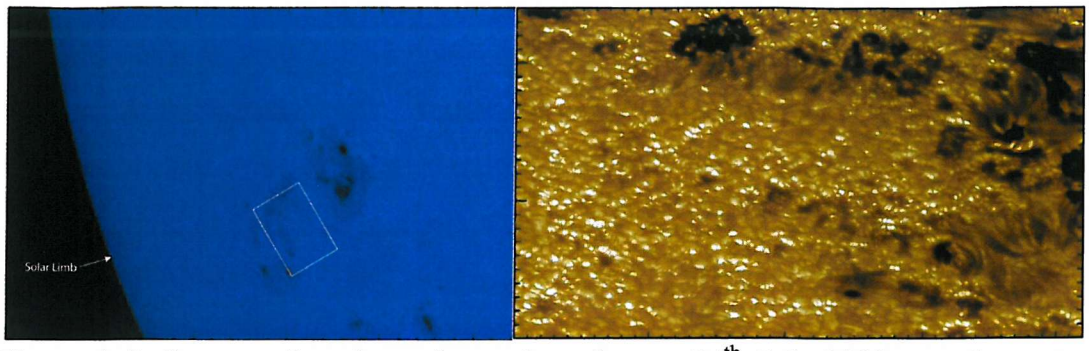


Figure 4. 8. Images of a solar active region taken on 24th. July 2002, near the eastern limb of the Sun, as recorded by TRACE (left) and a filtergram taken in 488 nm light by the Swedish 1-meter Solar Telescope (SST) on the island of La Palma. The active region is seen in the left-hand plot in which tickmarks are 10,000 km apart and the white box outlines the approximate SST field-of-view in the image shown on the right. TRACE has 10 times less spatial resolution than the SST and so faculae show up here as only vague bright patches surrounding the active regions in the left image. Only when looking at active regions towards the solar limb with 70 km spatial resolution of the SST do the three-dimensional aspects of the photosphere and faculae become apparent. In the right hand figure, tick marks are 1000 km apart and the limb is towards the top of the right hand of this image. The smallest resolvable features in the image are about 70 km in size. The structures in the dark sunspots in the upper central area of the image show distinct elevation above the dark "floor" of the sunspot. There are numerous bright faculae visible on the edges of granules that face towards the observer. [*Berger et al.*, 2003].

4. 9. 4. Photometric Facular Index.

The theory presented for the PSI can be adapted to generate a photometric facular index, for a facular/micropore contrast

$$C = \frac{I_F}{I_{Qs}} - 1 = \frac{\left(I_f - I_{Qs}\right)}{I_{Qs}} \tag{4.66}$$

we derive, for the same Eddington limb darkening function:

$$\Delta I_f = Q_0 \sum_f \alpha_f C \frac{\left(3\mu_f + 2\right)}{4} = Q_0 \sum_f \left(\frac{A_f}{A_{sh}}\right) C\mu_f \frac{\left(3\mu_f + 2\right)}{2} \tag{4.67}$$

where α_f is the disk facular filling factor, and A_f is the surface area of faculae.

At unit optical depth the surface temperature in faculae is about 150K higher than the quiet Sun, and thus T_f = 6200K. Again assuming a blackbody radiator, (4.66) yields a contrast c_f = C for faculae of:

$$c_f = \left(\frac{T_f}{T_{Qs}} \right)^4 - 1 \approx 0.103 \qquad (4.68)$$

At sunspot maximum the total area of faculae is roughly 10 times that of sunspots, thus $\sum_f A_f$ peaks at sunspot maximum at about $(1.5\times10^{-3})A_{SH}$. From equations (4.68), (4.67) and (4.63), ΔI_f / Q_o = 0.21%, which for Q_o = 1365 Wm^{-2} yields $\Delta I_f \approx 3$ Wm^{-2}. Thus these broad considerations predict that sunspots case a darkening of about $1Wm^{-2}$, whereas faculae cause a brightening of about 3 Wm^{-2}: together these cause a net solar cycle variation in total solar irradiance of about 2 Wm^{-2}, as has been observed over recent solar cycles. Note that the facular contrast is roughly one third of that of spots, but that they cover roughly 10 times the areas.

Ortiz et al. [2002] have provided a more precise algorithm for computing the contrast of small flux tubes (faculae and micropores), as a function of the field observed in a pixel of the Michelson Doppler Interferometer (MDI) (http://soi.stanford.edu) on the SoHO satellite. Contrasts of MDI pixels were evaluated at 676.8 nm, using the definition given in equation (4.65), relative to a field-free quiet sun intensity, corrected for limb darkening. The limb darkening function used is that derived by *Neckel and Labs* [1994].

The SoHO MDI instrument measures the continuum emission at 676.8 nm. *Neckel and Labs* give the coefficients for a limb darkening function L_D, a polynomial function of the disc parameter function μ: the nearest wavelengths in their table of coefficients are 669.400 nm and 700.875 nm and we here linearly interpolate to the values for 676.8 nm.

$$L_D = 0.3544 + 1.3472\mu - 1.9654\mu^2 + 2.5854\mu^3 - 1.8612\mu^4 + 0.54\mu^5 \qquad (4.69)$$

The resulting limb-darkening function, $L_D(\mu)$ is shown in figure 4.9, which is similar to, but more precise than the Eddington function that is also shown in the figure. Because L_D is a function of μ only, it can be calculated that its disc-average value is:

$$\langle L_D \rangle_D = 2\int_0^1 \langle L_D \rangle \mu d\mu \tag{4.70}$$

which equals 0.8478 and is shown as the dashed line in figure 4. 9.

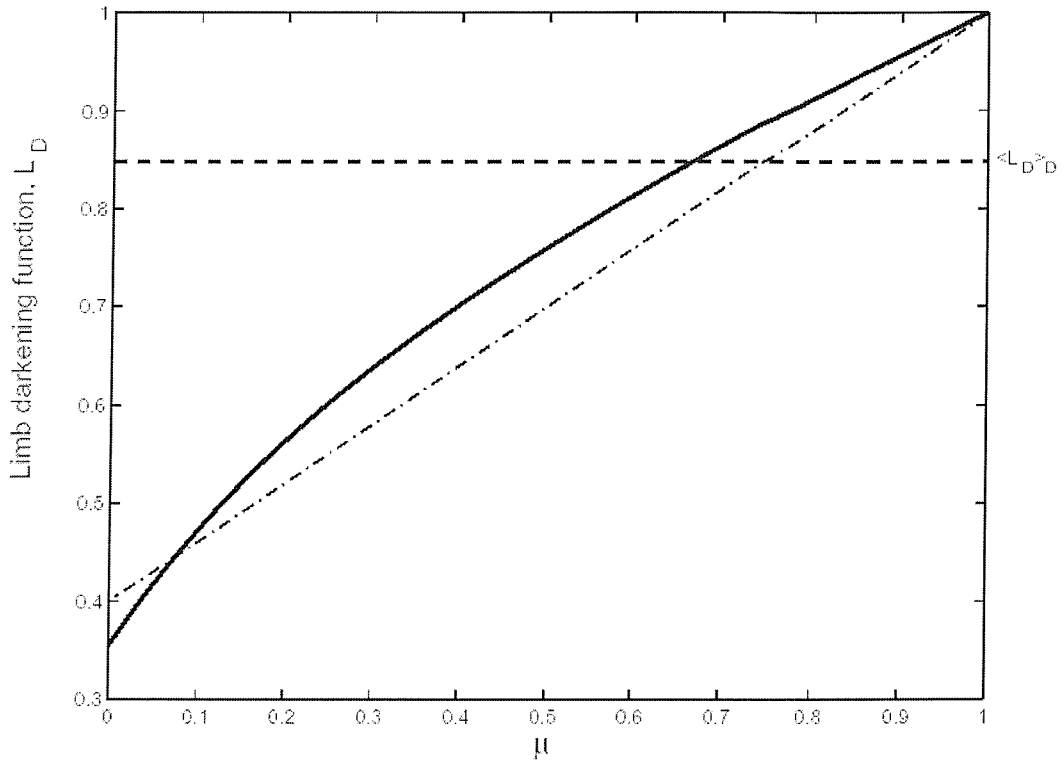


Figure 4. 9. (Thick solid line) the limb-darkening function, $L_D(\mu)$ for a wavelength of 676.8nm from the polynomial expression by *Neckel and Labs* [1994], as given in equation (4.69). The disc-average value is shown by the horizontal dashed line $\langle L_D \rangle_D$ = 0.8478. For comparison, the thin dot-dash line shows the Eddington limb darkening function used in the derivation of the photometric sunspot index.

Ortiz et al studied the contrasts as a function of (B_{MDI} /μ) and μ where B_{MDI} is the field detected by MDI in the line-of-sight direction. If we assume the field is radial, it has a magnitude of (B_{MDI} /μ) and plots like figure 4.10, shows that this yields very low scatter in the contrast data. The lines in figure 4.10 are *Ortiz et al.*'s polynomial fit to the data.

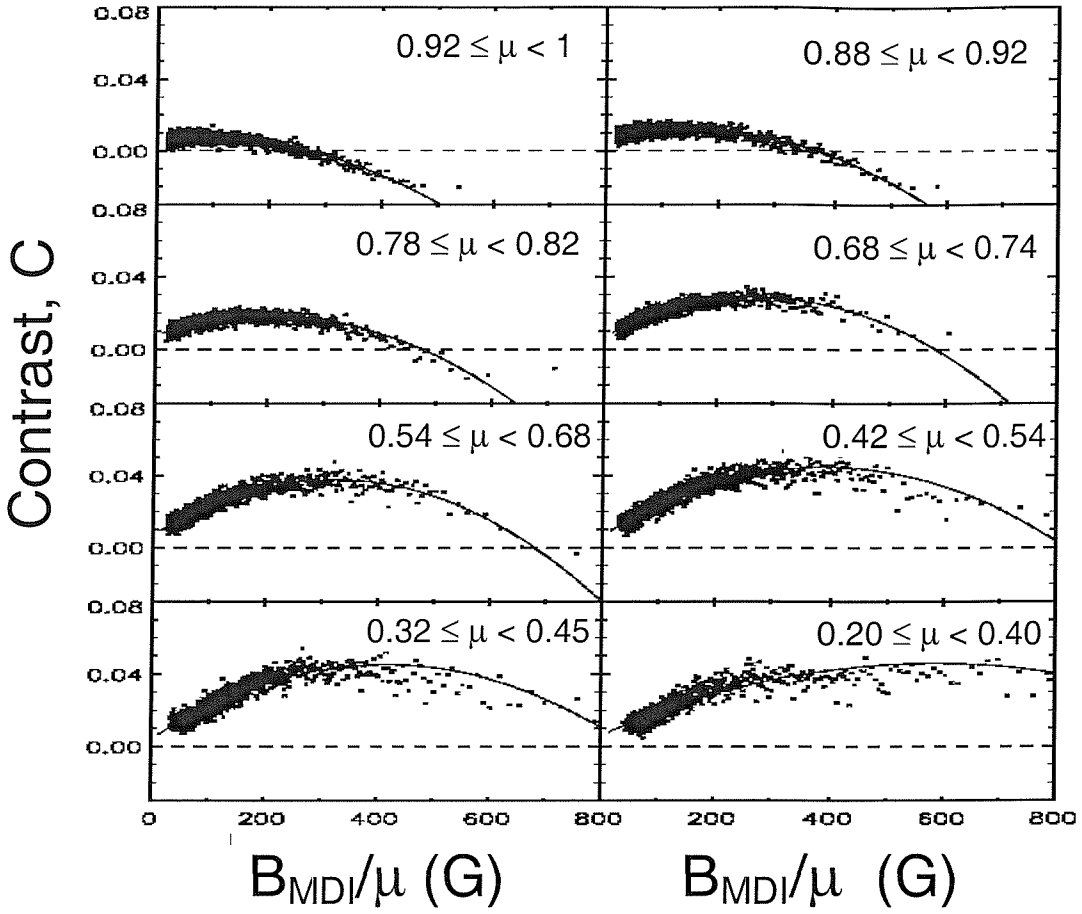


Figure 4. 10. Observations of the facular/micropore contrast C as observed at 676.8 nm by *Ortiz et al* [2002] and sorted as a function of the disk position parameter μ and the radial field component (B_{MDI} /μ), where B_{MDI} is the line-of sight field observed in an MDI pixel. The lines show the fits using the algorithm given in equation (4.71).

It should be noted that these data are specific to the pixel size of the MDI data (2″× 2″). This is because the facular flux tubes are not resolved and all faculae tend to have roughly the same field of $B_f \approx$ 1000G. This means that the B_{MDI} values are strongly dependent on the facular filling factor in a pixel, rather than the value of B_f.

Ortiz et al. [2002] derive the best-fit polynomial:

$$C\left(\left|\frac{B}{\mu}\right|,\mu\right)=\left(0.48+9.12\mu-8.50\mu^2\right)\times10^{-4}\times\left|\frac{B}{\mu}\right| \qquad (4.71)$$

$$+\left(0.06-2.00\mu-1.23\mu^2\right)\times10^{-6}\times\left|\frac{B}{\mu}\right|^2$$

$$+\left(0.63+3.90\mu+2.82\mu^2\right)\times10^{-10}\times\left|\frac{B}{\mu}\right|^3$$

which is plotted in figure 4. 11.

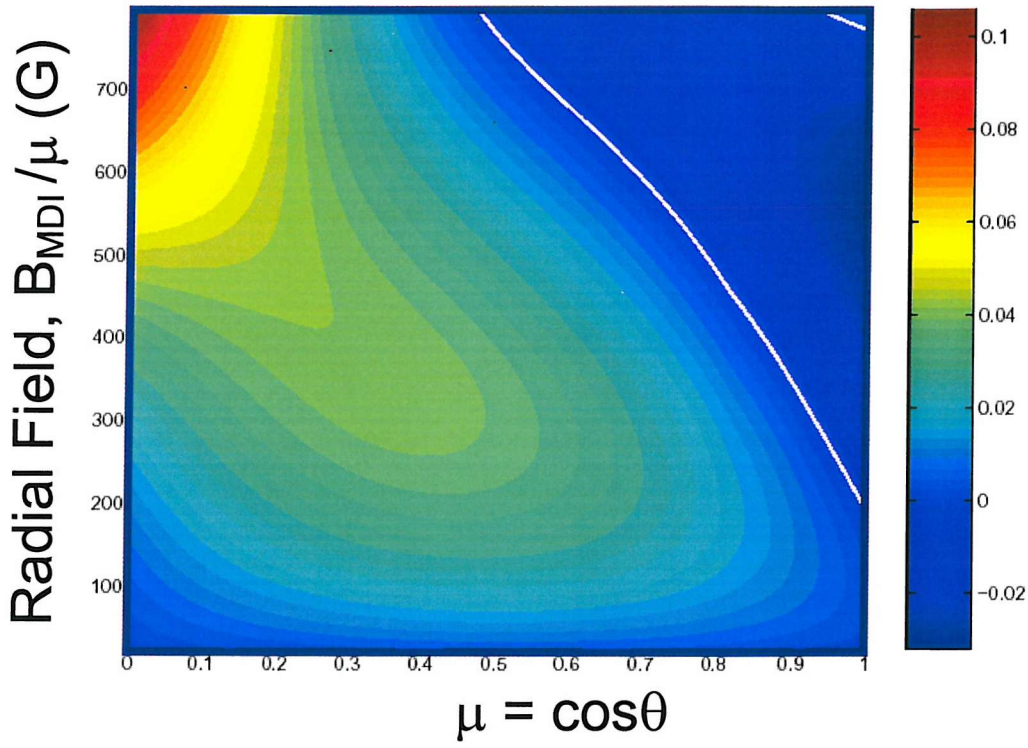


Figure 4. 11. Plot of the contrast of faculae and micropores at 676.8 nm, as a function of radial field strength |B/μ| and disk position μ, from the polynomial fit by *Ortiz et al.* [2002], as given by equation (4.71). Note that dark micropores are observed near the disk centre (μ near unity) and larger field values. Both contrasts and field values relate to pixels of the size of the MDI instrument.

Ortiz [2003] has demonstrated that these contrasts do not vary with the solar cycle.

Chapter 5: Intercalibration of Greenwich, USSR, Mount Wilson and USAF data sets.

5.1. Intercalibration of sunspot latitude measurements.

The intercalibration presented here is required because the Greenwich sunspot observations concluded in December 1976, and I wish to extend our data series using the observations made at the Mt. Wilson observatory. These observations were made in a compatible way to the Greenwich data and so can be considered a usable extension to the Greenwich data. However, the details of the intercalibrations required, vary from parameter to parameter. A complication is that data from Debrecken, Hungary for the period 1977-1981, have been used to continue the Greenwich observations and intercalibrate with the Mt. Wilson observations; however the Debrecken data series cannot be seen as a homogeneous extension of the Greenwich observations.

To verify the validity of continuing the annual sunspot latitude data series with the Mt. Wilson data, I investigated the crossover period between the two data series. It was found that there was excellent agreement between the two sites, with the highest correlation being found for the period 1967-1981. Figure 5.1 compares the average latitudes obtained from the Greenwich and Debrecken observatories and the Mt. Wilson data sets, for the periods 1962-1981 and 1967-1981. Although the correlation between the two data sets for both periods is extremely high ($r = 0.9982$, and $r = 0.9997$ respectively), it can be seen that in the early data a number of spots occur away from the line of best fit. The reason for the early Mt. Wilson data being slightly different to the simultaneous Greenwich data can be attributed to differences in calibration of the instrumentation and in the observational techniques employed at the two sites. As observations continued at Mt. Wilson, these differences were corrected until the results from the Mt. Wilson observatory were comparable to the Greenwich observations. Thus, we believe it to be valid that we continue the Greenwich data series with the Mt. Wilson data after 1981 using intercalibration derived from the

1967-1981 overlap period. Figures 3.1 and 3.2 show composite average latitude data sets, for both the northern and southern hemispheres respectively.

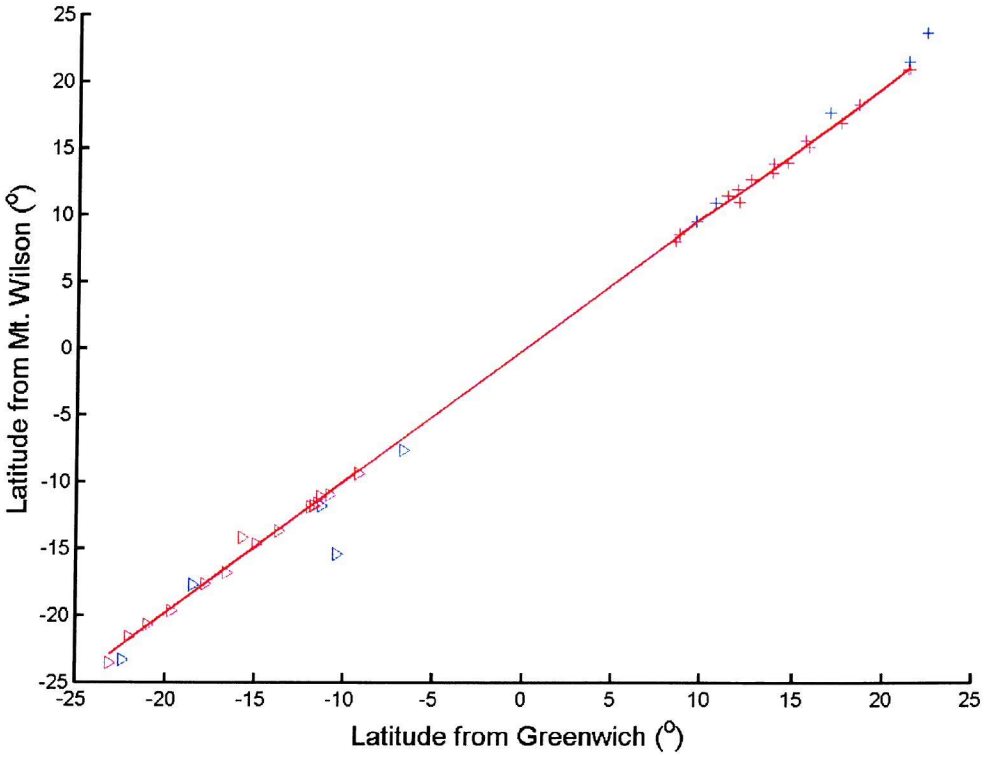


Figure 5.1. Comparison of the annual means of sunspot group latitudes, as observed at Greenwich and Mt. Wilson between 1962-1981 (blue) and 1967-1981 (red) in the northern (crosses) and southern (triangles) hemispheres. The solid line is the best-fit linear regression.

Calibration of the monthly values of sunspot latitude calculated in Chapter 3, are made in a similar fashion to the annual data series. Again the overlap period between the two data sets are investigated, and it is found that the two correlate well for the periods 1962-1981 and 1967-1981, r =0.9656 and 0.9861 respectively. As for the annual data, the correlation is lower for the period before 1967, and this is again attributed to calibration and observational differences between the two sites. Figure 5.2 compares the monthly average sunspot latitudes from the Greenwich and Mt. Wilson observatories, and it can be seen that the two data sets compare very well. Thus we can again use the Mt. Wilson latitude data series to continue the Greenwich monthly average latitude after 1981. Figure 3.17 gives the average of the northern and southern hemispheres latitudes respectively.

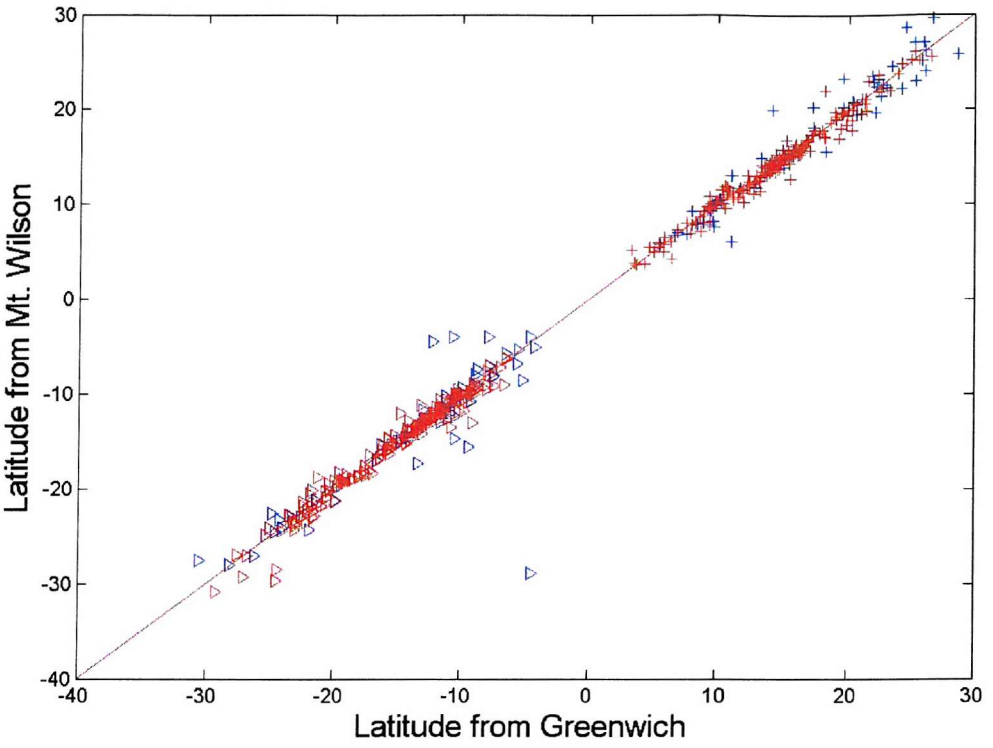


Figure 5.2. Comparison of the monthly means of sunspot group latitudes, as observed at Greenwich and Mt. Wilson between 1962-1981 (blue) and 1967-1981 (red) in the northern (crosses) and southern (triangles) hemispheres. The solid line is the best-fit linear regression.

5.2. Intercalibration of sunspot spread measurements.

As outlined in Chapter 3, I calculated annual sunspot spread from the Greenwich and Mt. Wilson for the duration of each data set. As for the latitude data, I investigated the overlap period between the Greenwich and Mt. Wilson data sets to validate the continuation of the Greenwich data with the Mt. Wilson data. Figure 5.3 shows that the data from the two observatories agree closely, with a correlation coefficient of 0.98, although there is a systematic tendency for the Mt. Wilson spread measurements to be slightly lower than those from Greenwich. This can be seen in the figure where the best-fit linear regression lies slightly below the dashed line, on which all points would sit if the two data sets were identical.

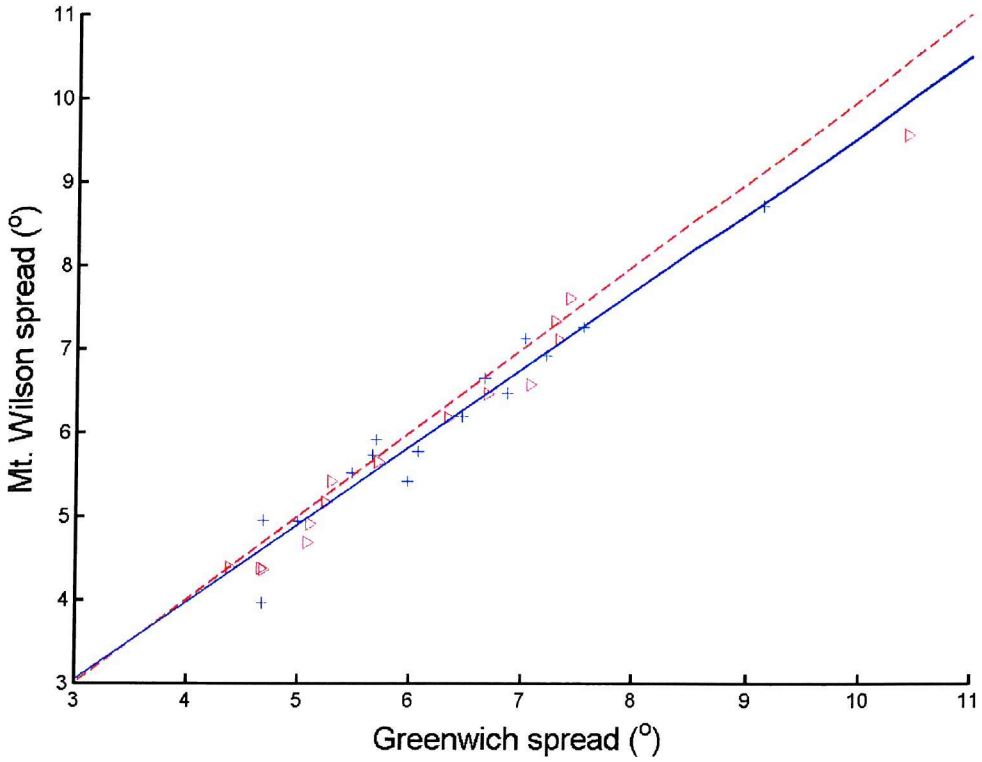


Figure 5.3. Comparison of the annual standard deviation of the sunspot group latitudes, as observed at Greenwich and Mt. Wilson between 1967 and 1981, in the northern (crosses) and southern (triangles) hemispheres. The solid line is the best-fit linear regression. The dashed line represents the ideal match between the two data sets.

To account for this 3% discrepancy, Mt. Wilson spread values can be multiplied by 1.03 to make them compatible with the Greenwich spread data. This good agreement between the two data sets allows us to append the Greenwich data set with the Mt. Wilson data set and obtain a homogeneous data set of sunspot spread for the period 1874-1999, which can be seen in figures 3.6 and 3.7.

Later in this chapter (section 5.3) I will confirm an often-quoted calibration factor of 1.4 by which Mt. Wilson total sunspot group area must be multiplied to make them consistent with the Greenwich data. If all sunspot groups were recorded in both data sets, but all groups scaled up by 39% in the Greenwich data, then the spread would be the same in both data sets. The fact that the Mt. Wilson spread is smaller shows that some of the 1.39 factor in area arises because some groups registered in the Greenwich data are omitted in the Mt. Wilson data. There appears to be no systematic

bias in the loss of these missing groups because the mean latitude is not affected (figure 5.1 and 5.2)

The scatter plot for the monthly spread is shown in figure 5.4. Again it was found that the Mt. Wilson observations gave a systematic tendency to under estimate the values of the sunspot spread, so the Mt. Wilson monthly values of sunspot spread were also multiplied by 1.03 to make them compatible with the Greenwich observations. The Mt. Wilson data series is then appended to the Greenwich data series, to get a uniform monthly data sequence of sunspot spread, spanning from 1874 – 1999. This full data sequence of monthly sunspot spread can be seen in figure 3.18 for the northern and southern hemispheres combined.

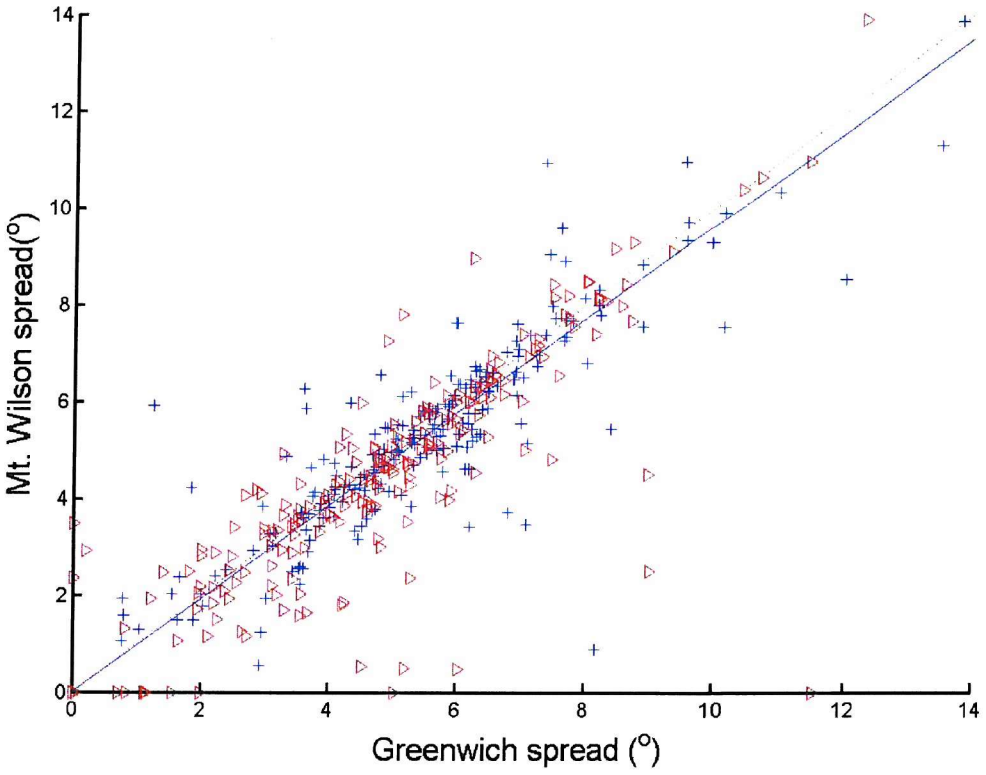


Figure 5.4. Comparison of the monthly standard deviation of the sunspot group latitudes, as observed at Greenwich and Mt. Wilson between 1967 and 1981, in the northern (crosses) and southern (triangles) hemispheres. The solid line is the best-fit linear regression.

Note, however, there is considerable scatter in figure 5.4, and this arises because for 8 months in the 14 years, no spots were recorded in the Mt. Wilson observations, whilst spots were observed at Greenwich (and/or its back up stations). In only 2 months were sunspot groups observed only at Mt. Wilson.

5.3: Intercalibration of sunspot area measurements.

As outlined in Chapter 2, sunspot area measurements were made at the Greenwich observatory, from 1874 to 1976. Observations from the Debrecken observatory have been appended to the end of the Greenwich data, so that a 'Greenwich' sunspot data series can be continued to December 1981. At the same time (December 1981) the USAF/Mt. Wilson commenced publication of sunspot area measurements, and these observations were used to keep a continuous data series of sunspot area from 1874 to the present day.

Unlike the sunspot latitude data, the group area was not recorded at Mt. Wilson until 1981, when recordings commenced under contract with the USAF. However in many datasets labelled 'Mt. Wilson' there are area data for 1976 – 1981, but on closer inspection these records appear to be identical to the 1976 –1981 'Greenwich' area data, implying they are actually the Debrecken sunspot area observations. This has led to a great deal of confusion, trying to assess when exactly data series are valid for, and what are the 'true' values for sunspot parameters that can be used to intercalibrate the Mt. Wilson and Greenwich data. Similarly, observations of sunspot group position, like area, only started in December 1981, carried out by the USAF in conjunction with the Mt. Wilson observatory. Intercalibration of these data sets is straightforward and systematic errors and problems have been found in the area and position data sets after 1976. There has been no real explanation as to why there are differences in the data sets, but I shall here speculate on what I believe to be the most likely origin of these errors. Since there was no real overlap between the Greenwich and USAF/Mt. Wilson observations, so as to compare and calibrate the instruments at the two sites, Debrecken observations was used by the USAF/Mt. Wilson observers to calibrate their procedures, under the assumption that the Debrecken observations were a close match to the Greenwich observations. The early area measurements were then matched to the Debrecken observations, and calibrated so that where the observations overlapped the area measurements matched. Under this assumption that the USAF/Mt. Wilson measurements were a good match for the Debrecken data, the Debrecken data was appended to the start of the USAF/Mt. Wilson data series, thus giving a data set back to 1976. Unfortunately this has lead to

problems. I do not believe that the data contained within the Debrecken data series is a good match to the Greenwich data for area and position.

Monthly sunspot area measurements from Greenwich showed the expected eleven-year solar cycle variation, but have also found to have steadily increased at solar maximum since 1900, with a peak being found in solar cycle 19 around 1959, and these area measurements are found to vary in a similar fashion and correlate well with sunspot number (as will be shown in Chapter 6). The first indication of a problem with the area measurements from the Debrecken observatory is that the Sunspot number/Area relationship, which is strong for 1874-1976, gets weaker after 1976, with the area measurements now being consistently smaller from both the Debrecken and USAF/Mt. Wilson observatories. On closer inspection it was found that the sunspot areas after 1976 were around 30% smaller than the earlier Greenwich observations, although this could not be verified since there was no real overlap between the data sets. Dr. Hathaway of the NASA solar data centre (see www.nasa.gov) has reached a similar conclusion. I have been able to confirm this finding independently with the aid of the previously unexploited USSR sunspot data records, which cover the period 1968-1991. The area measurements from the USSR observatory show a good match to the overlapping Greenwich observations (1968-1976), but diverge after 1976, with the USSR data being found to be 1.39 times larger than the Debrecken and USAF/Mt. Wilson sunspot area measurements. This is independent evidence for the need to revaluate the sunspot area measurements after 1976 made at the Debrecken and USAF/Mt. Wilson observatories.

On the basis of this evidence I have rescaled the my monthly sunspot area measurements after 1976, multiplying the uncorrected values by 1.39 so that the monthly sunspot area measurements are now consistent with the Greenwich observations, thus a continuous and reliable data set of monthly sunspot area is available from 1874 to the present day. It shall be this corrected data set that I shall use in my further studies, and I believe that with the points outlined above I can use this data series with confidence in further studies.

Care must be taken because the Debrecken and Mt. Wilson data have already been corrected by a factor of 1.4 in some series stored in data centres, but not in others. If it is unclear whether or not the scaling factor has been applied it is important to test against the USSR data.

5. 4. Long-term variation and intercalibration of sunspot μ values.

The value μ defines the position of magnetic features on the solar surface, and lines of constant μ form a series of concentric rings around the centre of the solar disc. At the centre μ = 1, and at the limb μ = 0. As the feature rotates across the surface of the Sun the μ value will vary (see in figure 5.5). If a feature appears at a high latitude then its μ value will remain low and will vary little, since the feature will exist only near the limb, if the feature appears at the solar equator it's μ value will vary largely from 0 to 1 and back again as it moves across the visible disc.

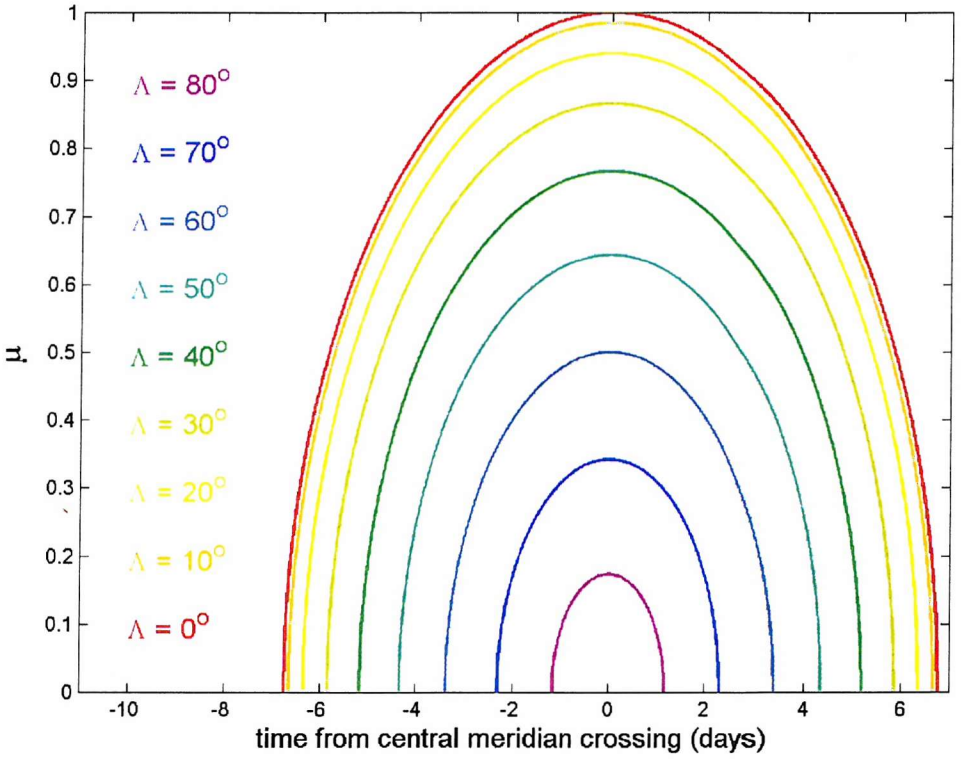


Figure 5.5. Variation of sunspot μ with time and latitude.

The value of μ can be calculated in a number of different ways, either from the latitude and longitude of the feature, or using the measured value r contained within the dataset (the distance from the centre of the solar disc in solar radii), or the heliocentric angle θ. These are related by the simple geometry shown in figure 5.6

$$\mu = \cos\theta \tag{5.1}$$

$$\sin\theta = \frac{r}{R} \tag{5.2}$$

$$\mu^2 = \cos^2\theta = 1 - \sin^2\theta \tag{5.3}$$

$$\mu^2 = 1 - \frac{r^2}{R^2} \tag{5.4}$$

$$\mu = \sqrt{1 - \frac{r^2}{R^2}} \tag{5.5}$$

If the r value is not available or was not recorded in the data series then it is possible to calculate a μ value for a feature from its solar latitude (Λ) and the longitude (L).

$$\sin\theta = \frac{Q_s}{R_s} = R \tag{5.6}$$

$$\cos\theta = \sqrt{1 - \sin^2\theta} = \sqrt{1 - R^2} \tag{5.7}$$

$$\cos\theta = \mu = \frac{OQ}{OS} = \frac{OQ}{R_s} \tag{5.8}$$

$$\frac{OP}{R_s} = \cos\Lambda \tag{5.9}$$

$$\frac{OQ}{OP} = \cos L \tag{5.10}$$

$$\mu = \cos\theta = \frac{OQ}{R_s} = \cos\Lambda \cos L \tag{5.11}$$

Values of latitude, longitude and r were recorded for each sunspot groups in the data from the royal Greenwich observatory, USAF/Mt. Wilson observatory and various other observatories, since 1874, and so it is possible to calculate the μ value for individual groups, and to monitor how the positions of features has changed since 1874.

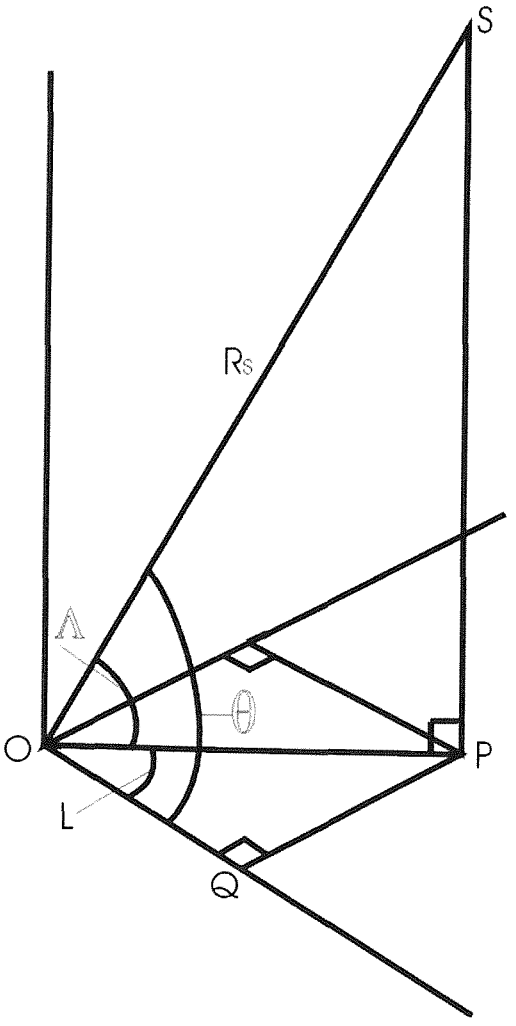


Figure 5.6. Three-dimensional representation of sunspot position on the solar surface. O: Solar center, S: position of sunspot, R_S: radius of sun, L: longitudinal angle, Λ: latitudinal angle, θ: positional angle.

Chapter 3 discussed how the latitudes at which sunspots appear have varied over the past 100 years, and this means that the values of μ will also display a long-term change, since μ is partially dependent upon latitude. Monthly values of μ have been calculated from the Greenwich and USAF/Mt. Wilson observations, and this can be seen in figure 5.7, and it can be seen that the average value of μ has decreased over the past 100 years consistent with the rise in average latitude. The monthly μ values also display a solar cycle variation consistent with the butterfly latitude variations, with μ being smaller at the beginning of the solar, increasing through the solar cycle, with a maximum being reached at the end of the solar cycle, before it falls at the start of the next solar cycle. In addition to the Greenwich observations up to 1976, figure 5.7 shows the Debrecken and USAF/Mount Wilson observations, which continue to the present day. It can be seen that during the period 1976-1981, the μ values do not follow the usual solar cycle variation that has been evident during all the previous solar cycles. The reason for this is unknown, but it does coincide with the change to the Debrecken observations and is obviously a cause for concern, and to verify the results we investigated μ values from other observatories. Several observations were looked at, including the Taipei, Rome, Yunnan, and USSR observatories. The results are very similar in each case but the USSR data series is the most valuable as it gives long and equal overlaps with both Greenwich and Mt. Wilson data. This data is plotted in the figure 5.7, in red. As it can be seen the USSR data shows that there is a solar cycle variation in μ during the period in question, the same variation has also been found in the Yunnan data. The reason why the overlap Greenwich/Debrecken and USAF/Mt. Wilson datasets show a μ variation that is so different is not obvious, but is likely to have arisen from a misinterpretation of the Debrecken data. When the actual USAF/Mt. Wilson observations commenced in 1981, the values or r (and hence μ) resumed the normal solar cycle variation, as seen in the Greenwich data before 1976. Given that the latitude data from Debrecken are largely consistent with the prior Greenwich data and the subsequent Mt. Wilson data (see section 5.1) the error appears to have arisen in the longitude or r data field. For the observations outside the suspect 1976 - 1981 period, the USSR, Greenwich and USAF/Mt. Wilson observations all correlate extremely well. I have therefore constructed a composite which uses the monthly USSR μ data for the period 1977-1985.

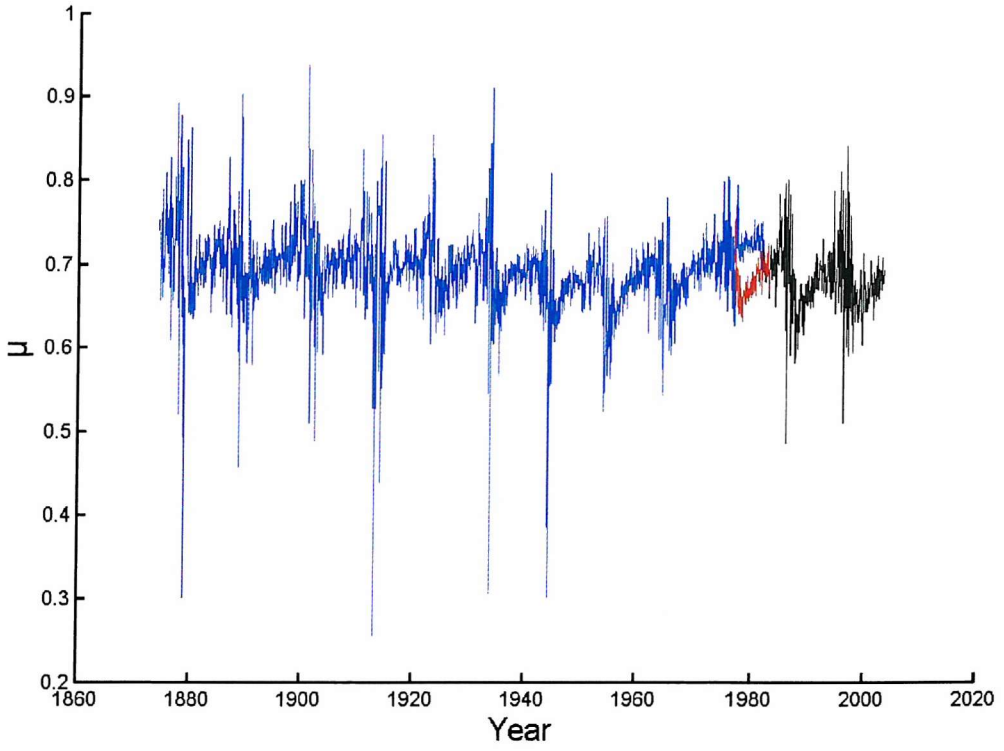


Figure 5.7. Monthly mean variation in sunspot μ values for the period 1874-2003. Blue = Greenwich/Debrecken, red = USSR and black = Mt. Wilson/USAF.

5.5. Intercalibration of modelled facular brightening.

As has been stated in sections 5.4 and 5.3, continuous measurements of sunspot area and position have not been made at a single observatory and thus intercalibration of the data sets is needed. In chapter 7 of this thesis I shall use both sunspot area and μ values from the Greenwich, Mt. Wilson and USSR observatories to model the long-term variation in facular brightening, but again some intercalibration of the results will be needed, due to the effects of the discrepancies between the measurements from each of these observatories as outlined in the previous sections. The data for the following periods from the following observatories have been used to develop the facular brightening model; 1874-1976 Greenwich, 1977-1985 USSR and 1986-2004 Mt. Wilson observations (N.B. I use the USSR to 1985, so data spans minima to minima to limit any problems that may occur introducing a data set midway through a solar cycle). This means that for solar cycles 12-20 the Greenwich

observations shall be used, cycle 21 the USSR observations shall be used, and for cycles 22 onwards the Mt. Wilson observations. To check the intercalibrations in μ and A_S, I here assess the inferred facular brightening with and without the intercalibrations discussed earlier. Figure 5.8 shows the scatter plot of modelled facular brightening from the Greenwich and USSR observatories against sunspot number. Sunspot number is used as it is well correlated with sunspot area and is available in a highly homogeneous form throughout the relevant interval. As can be seen for the period before 1977 (when the observations are believed to be correct) the two data sets match up well to sunspot number, but for the period after 1977 the facular brightening calculated from the Greenwich/Debrecken data series is found to be lower than that calculated from the USSR data set. Figure 5.9 shows the same scatter plot for the modelled facular brightening from the Mt. Wilson and USSR data series, and again it can be seen that for the period where the area and μ positions are in doubt, the Mt. Wilson modelled brightening is lower. Figure 5.10 shows the modelled facular brightening from the Mt. Wilson observatory, for which the area has been scaled by a factor of 1.39, to take into account the discrepancies between the Greenwich/Debrecken and Mt. Wilson area measurements. As can be seen, once the data has been scaled it shows a similar linear-variation as the USSR data, but it must be noted that there is still a good deal of scatter, and this is due to the fact that μ is a factor in the calculation as well as the sunspot area, and so adds to the scatter. Thus although the sunspot area can be corrected, the μ values for the period 1977 –1981 from Debrecken remain incorrect and lead to errors in the calculation.

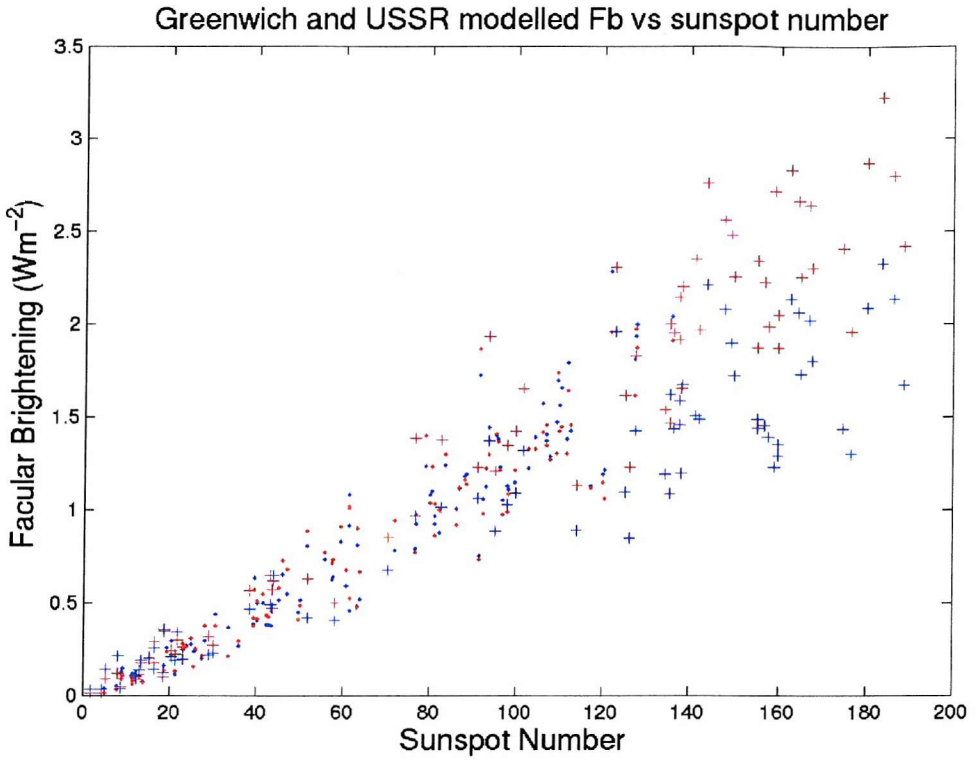


Figure 5.8. Scatter plot of monthly means of modeled facular brightening against sunspot number. Red: USSR data, blue: Greenwich data, crosses: 1977-1981, dots: 1968-1976. No intercalibration factors have been applied.

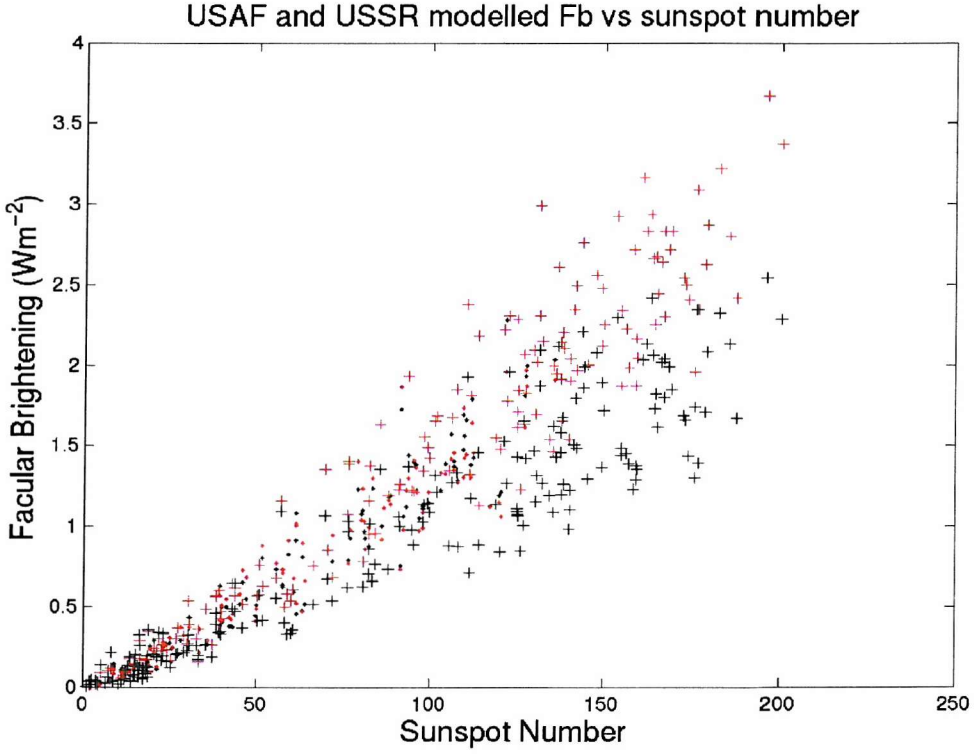


Figure 5.9. Scatter plot of monthly means of modeled facular brightening against sunspot number. Red: USSR data, black: Mt. Wilson data, crosses: 1977-1991, dots: 1968-1976. No intercalibration factors have been applied.

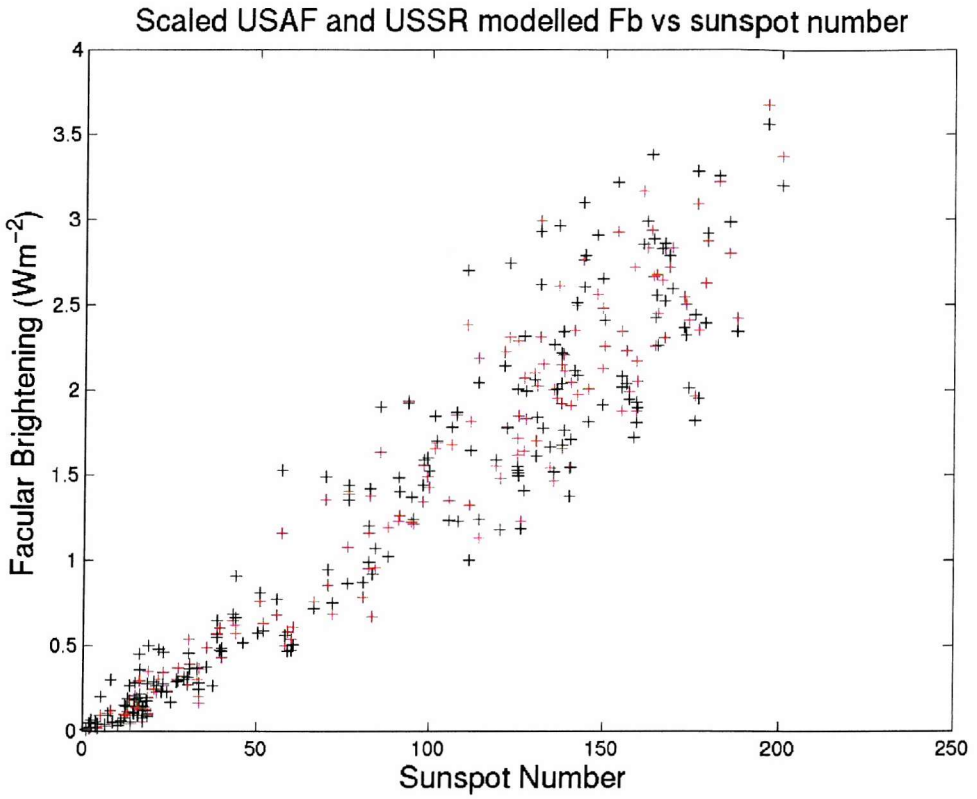


Figure 5.10. Scatter plot of monthly means of modeled facular brightening against sunspot number for the period 1977-1991. Red: USSR data, black: Corrected Mt. Wilson data.

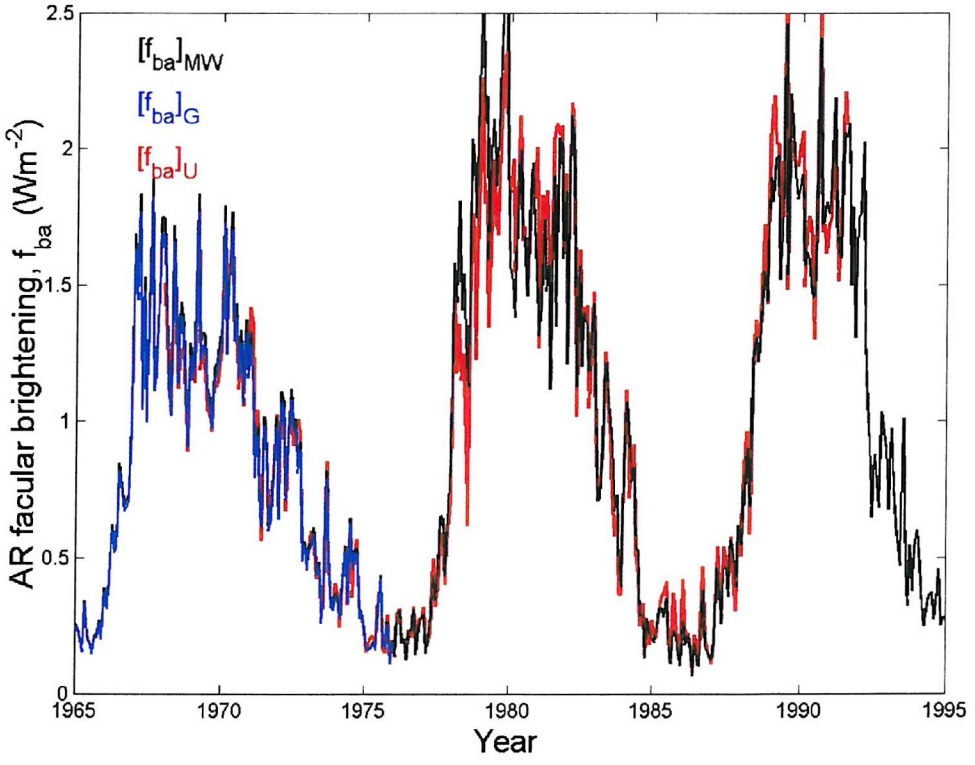


Figure 5.11. Monthly facular brightening estimates derived from sunspot group data from three different sources. In blue are data from Greenwich $|f_{ab}|_G$, in red are data from the former Soviet Union $|f_{ab}|_U$, and in black the USAF/Mount Wilson data (to which the inter-calibration parameters have been applied), $|f_{ab}|_{MW}$. Scatter plots are presented in figures 5.12-5.15.

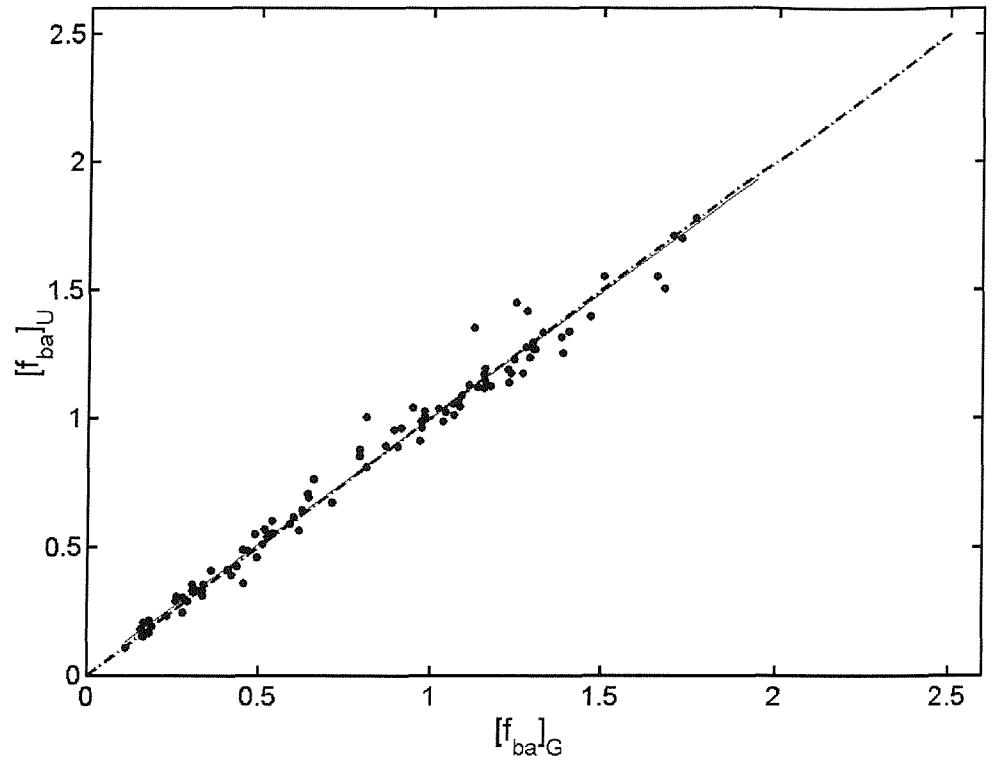


Figure 5.12. Monthly averages of the facular brightening derived from Greenwich sunspot data, $|f_{ab}|_G$, plotted against simultaneous values from the former Soviet Union $|f_{ab}|_U$. The solid line is the best-fit linear regression and the dashed is the ideal fit with $|f_{ab}|_U = |f_{ab}|_G$.

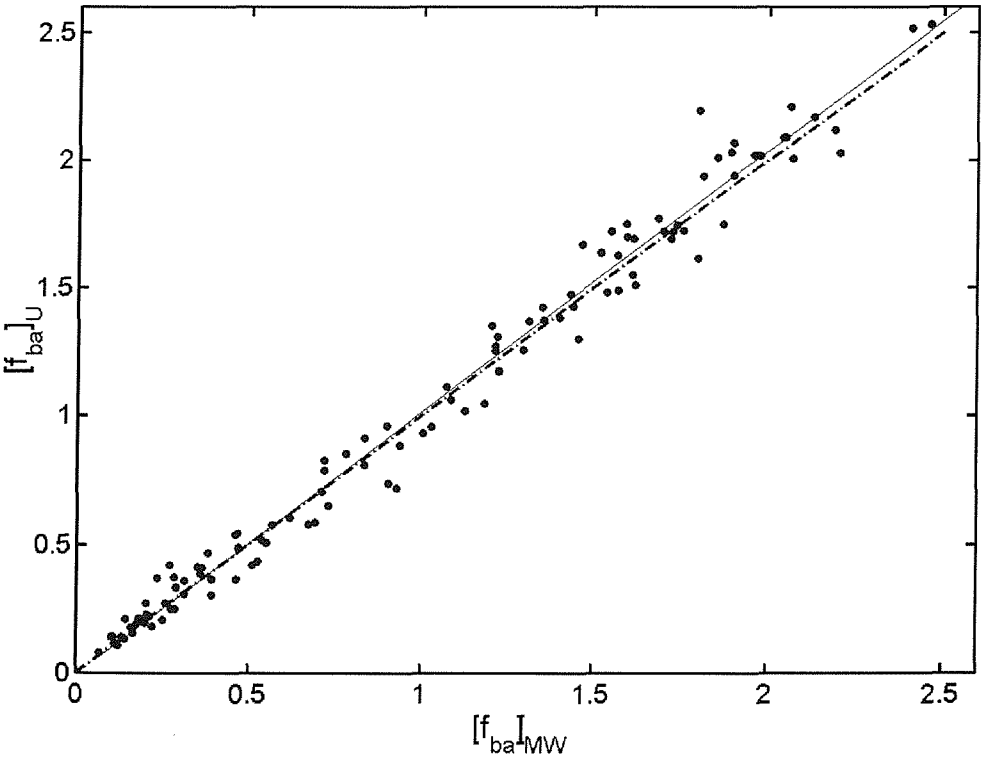


Figure 5.13. The same as figure 5.12, for USSR and Mt Wilson data, $|f_{ab}|_U$ and $|f_{ab}|_{MW}$, respectively, for after 1982 and after.

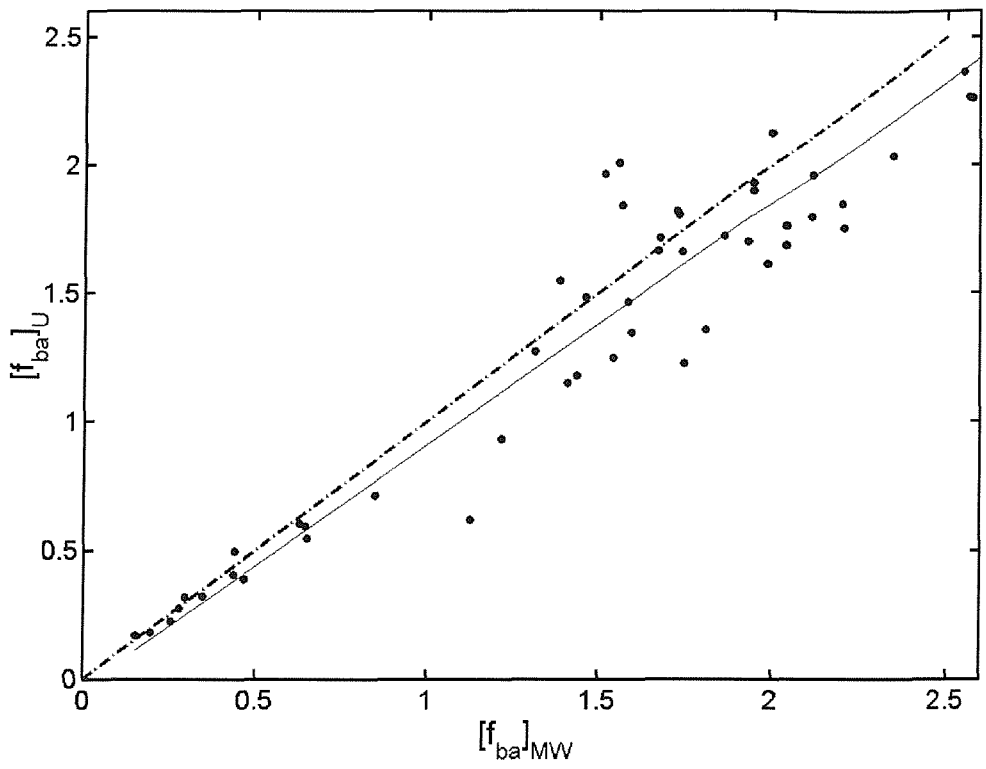


Figure 5.14. The same as figure 5.12, for USSR and Mt Wilson data, $|f_{ab}|_U$ and $|f_{ab}|_{MW}$, respectively, for 1976-1981.

The composite of the active region facular brightening derived from the monthly sunspot data is confirmed by figures 5.11-5.15. Figure 5.11 shows that the Greenwich and USSR data (blue and red lines, $|f_{ab}|_G$ and $|f_{ab}|_U$) agree very well up to the cessation of the Greenwich data in December 1976. This is confirmed by the scatter plot shown in figure 5.12, which is for the interval 1968 (the start of the USSR data) to 1976 (the end of the Greenwich data). The solid line is the best regression fit and the dashed line is the ideal line for which $|f_{ab}|_U = |f_{ab}|_G$. It can be seen that the agreement is extremely close indeed.

Similarly the agreement between the USSR data and the calibrated Mt Wilson data ($|f_{ab}|_U$ and $|f_{ab}|_{MW}$, red and black lines in figure 5.11, respectively) is extremely close for 1982 and after. This is best seen in Figure 5.13, which is a scatter plot for 1982 to 1991 (the end of the USSR data).

However, if we use the Mt Wilson data for 1976 to 1981 we do not see as good agreement as in figures 5.12 and 5.13. There is considerable scatter and the best-

fit regression is not close to the ideal dashed line. This is despite the fact that the Mt Wilson data have been corrected in the same manner as the data in figure 5.13. The large scatter in part arises from the fact that the μ values are not reliable in the Mt Wilson data for this interval Figures 5.11-5.13 show that the intercalibrations used here have ensured that the active region facular brightening data sequence, derived from the different sunspot datasets form a homogeneous series.

5. 5. Conclusions.

It is clear from the work in this chapter that an in-depth and thorough study of these long term data sets needs to be carried out to understand exactly what has occurred at various observatories, so that a reliable and continuous sunspot data series can be obtained. The corrections and intercalibrations required depend on the parameter. For average sunspot latitude the two datasets are very similar and an intercalibration factor of unity is required. However, the standard deviation of the latitudes requires 3% correction showing that not all the spots recorded in the Greenwich data would have been registered in the Mt. Wilson data. The biggest correction needed is a factor of 39% that must be applied to the sunspot group area. It is also evident that the USSR observations contain the only reliable and consistent μ values for the period 1977 – 1981.

Chapter 6: Investigation of the correlation between total solar irradiance and the coronal source flux, and the implications for the long-term variations of facular brightening and the quiet Sun.

6.1. Chapter overview.

Total solar irradiance has been measured using various instruments onboard satellites for the past two and a half solar cycles, but to quantify any effect that the Sun could be having upon global warming, a long-term and robust reconstruction of irradiance is needed. A strong correlation between open solar flux and TSI has been noted by *Lockwood and* Stamper [1999] and *Lockwood* [2002a] and is often implicitly assumed in paleoclimate studies using cosmogenic isotopes. In this chapter, I shall investigate the implications of this correlation for facular brightening. The TSI is reconstructed past the end of the open flux observations back to 1700, using the model of the open solar flux, F_S created by *Solanki et al.* [2000].

Monthly Photometric Sunspot Index is reconstructed using sunspot area for the period 1874-2000, and sunspot number for the period 1702-2000. Annual PSI is reconstructed using sunspot number for the period 1700-2000. Utilising the reconstructions of TSI and PSI, monthly and annual reconstructions of facular brightening are made. These reconstructions are used to quantify the various factors that are needed to account for the long-term increases in TSI that are implied by the open flux-TSI correlation.

6. 2.1. Total Solar Irradiance and the Coronal Source flux

As discussed in section 2.3.1, *Lockwood et al* [1999a] developed a method for calculating the coronal source flux, F_S, from yearly observations of the geomagnetic aa index. Monthly values can be obtained using a 12-month data window, which was

progressed by one month at a time. These F_s values have some inherent smoothing, with only every twelfth value being truly independent. Figure 3.19 shows the variation of these monthly F_s values for the period 1868-2001.

Lockwood and Stamper [1999] reported a strong relationship between annual values of F_s and total solar irradiance (TSI), with a correlation coefficient of r = 0.74. Using monthly means of TSI from the PMOD composite data sequence (see chapter 2), an investigation was undertaken into the correlation between TSI and monthly F_s estimates derived from the aa index. Figure 6.1, shows the scatter plot and the best-fit linear regression between F_s and TSI for the period 1982-2001. The best correlation coefficient is r = 0.76 which is found for the open flux lagging behind the TSI by 5 ± 4 months. The best-fit regression line is $I_{Fs} = sF_s + c$, and has slope $s = 5.672 \times 10^{-15}$ $Wm^{-2}Wb^{-1}$ and intercept c =1363.4 Wm^{-2}. The correlation drops for the period before 1982, and the reason for this is unclear if this is due to a physical or instrumental effect. The correlation is significant at the 99% level. The correlations and their significance of TSI with various open flux estimates and the cosmic ray fluxes have been tabulated by *Lockwood* [2002a].

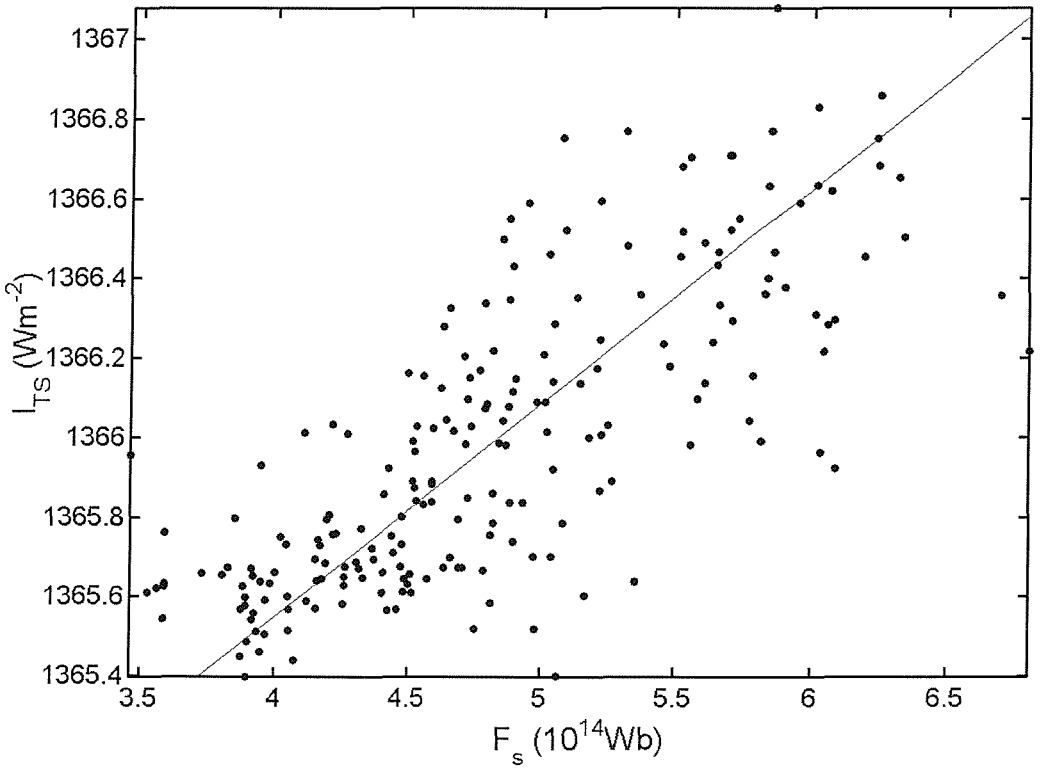


Figure 6.1. Scatter plot of the monthly values of F_s against TSI.

The best-fit linear regression shown in figure 6.1 is used to scale the monthly F_s data sequence to give the variation of I_{Fs}. The results can be seen in figure 6.2 for the period 1982-2000 in red, compared with the satellite measurements of TSI, I_{TS} (in blue). It can be seen that the variations over the solar cycle are very similar, but there are small-scale differences on monthly timescales. It is also interesting to note that as in the PMOD composite of TSI irradiance, where the minimum around 1997 is slightly lower than that around 1986, the same feature is seen in F_s data set. In contrast, the ACRIM composite shows the opposite trend in the minima (see figure 2.1).

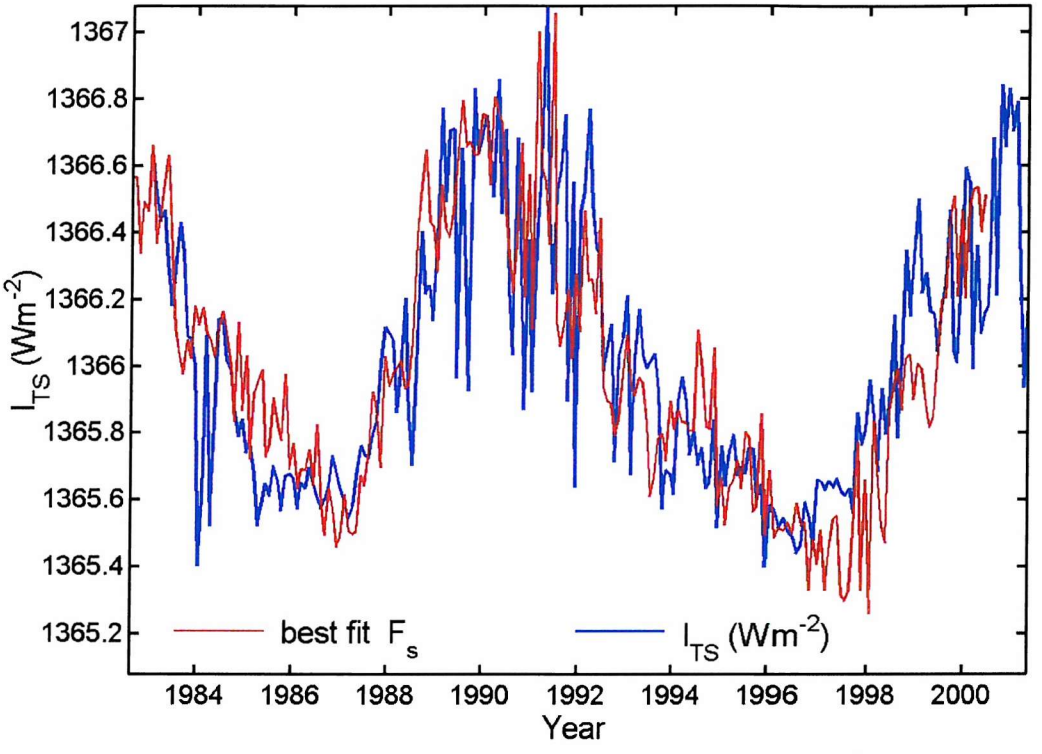


Figure 6.2. Monthly mean values of reconstructed and observed TSI for the period 1982-2000.

It has been found that recent irradiance changes are well explained by changes in the total flux and spatial distribution of the magnetic flux threading the photosphere [*Fligge et al.*, 1998; *Solanki et al.*, 2002]. The Earthward component of this field, B, has been measured by ground and space-based magnetograms, and a data sequence covering more than 2 solar cycles and with 1.4" × 1.4" resolution has been obtained from the Kitt Peak observatory. If I assume that this field is normal to the solar surface, it has magnitude $|B/\mu|$ where $\mu = \cos\theta$ and θ is the heliocentric angle of the

flux tube. I employ $|B|_{av}$, the disc average of $|B|$, the absolute value of the Earthward photospheric field component, as observed at Kitt Peak. The best correlation coefficient between the Kitt Peak average field magnitude $|B|_{av}$ and I_{TS} is 0.79 at a lag of 2 months, and figure 6.3 shows monthly means of the observed TSI from satellite measurements, I_{TS}, in blue, with the best fit of monthly averages of $|B|_{av}$, I_B, in red. Using a Fischer-Z test it was found that the difference between this correlation and that for I_{TS} and F_s (r = 0.76) is significant at the 75% level. Furthermore, the I_{TS}-$|B|_{av}$ correlation remains relatively good prior to 1982. From this I can infer that if F_s is a valid proxy for TSI, it is likely to be because of the better correlation with the photospheric field. This requires that the ratio F_s/F_p be relatively constant, where F_p is the total (unsigned) flux threading the photosphere.

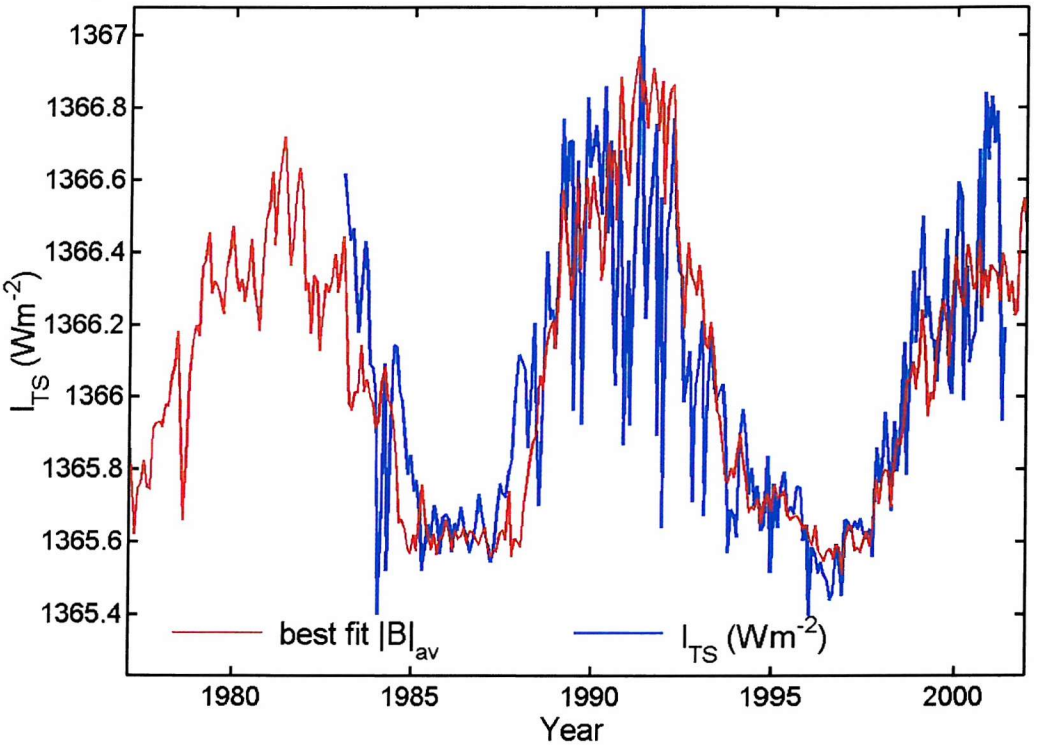


Figure 6.3. Monthly means of the observed and reconstructed TSI from Kitt Peak magnetograms.

Figure 6.4 shows the full data sequence of 12-point running means of the monthly I_{Fs} estimates, in blue, derived from the regression shown in figure 6.1, for 1868-2000. This reconstruction is similar to the annual values derived by *Lean* [2000a] (shown in red) from a combination of sunspot number, R, and it's 11-year running mean, R_{11} (acting as proxies for TSI variations on solar cycle and century

length timescales, respectively). The black line shows the TSI observations. The correlation between I_{Fs} and the *Lean* reconstruction is r = 0.84 which is significant at the 99.0% level.

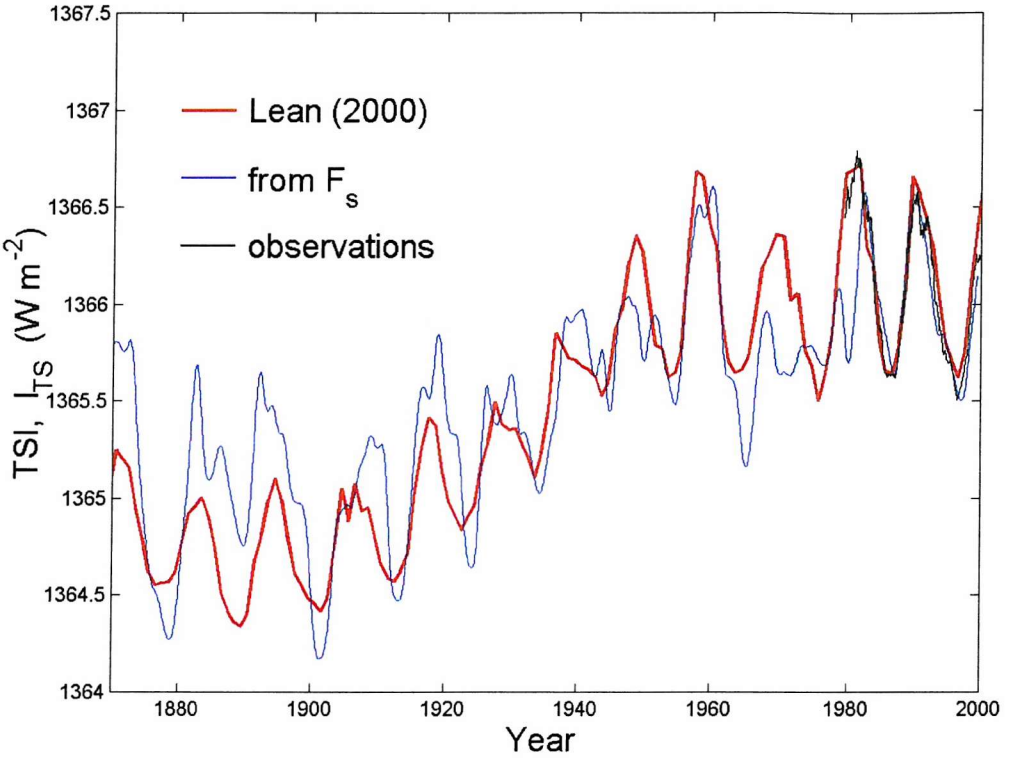


Figure 6. 4. Twelve-month running means of TSI calculated from monthly F_s data and the *Lean* TSI reconstruction.

Lockwood [2001] noted that the 11-year running means of sunspot number and open flux, R_{11} and $[F_s]_{11}$, varied in a similar fashion and so it is not surprising that these two TSI reconstructions have very similar patterns of long-term drift. The solar cycle oscillation amplitude is smaller at earlier dates in the *Lean* reconstruction, whereas it has roughly constant amplitude in the reconstruction based on open flux, (termed I_{Fs}). However, the amplitudes of the long-term drifts of the two reconstructions are also very similar. This is significant because *Lean* calibrated the amplitude of the long-term variation by comparison with non-cyclic sun-like stars, presumed to be in a state of equivalence to the Maunder-minimum Sun. Whereas, the drift in the reconstruction based on open flux, I_{Fs}, arises from the correlation over the last 2 solar cycles. The agreement between our TSI reconstruction and that by *Lean*

implies that the correlation between F_s and TSI holds on 100-year timescales as well as the 10-year timescales covered by the TSI observations.

The correlation between TSI and F_s, if valid on 100 year timescales, would allow us to reconstruct the variation in TSI back to 1700 using the model of the variation of F_s by *Solanki et al.* [2000]. This model is based upon the continuity equation for the total open solar flux and uses a complex function of annual sunspot number, R, to estimate the rate of emergence of new open solar flux. *Solanki et al.* [2000] ran this model forward in time, starting at an initial condition of zero open flux at the end of the maunder minimum. Here I use the variation by *Lockwood* [2001] who used the same model but ran it backwards in time, starting from the open flux seen in recent years (for example by the Ulysses satellite, IMF monitors or from solar magnetograms analysed by the PFSS method –see *Lockwood* [2004]). Figure 6.5 shows annual TSI values scaled from this modelled F_s variation, using the linear regression fit shown in figure 6.1. The F_s, and thus the inferred I_{TS}, is low at the start of the data sequence because this is at the end of the Maunder minimum when sunspot number, flux emergence rate and open flux were all reduced. Subsequently F_s and I_{TS} rose rapidly before falling to a second minimum between 1795-1820, corresponding to the Dalton minimum in sunspot numbers. After this minimum values again recovered rapidly before showing a smaller minimum around 1900 and rising to a peak at the maximum of cycle 19 (around 1955).

I shall adopt the notation that δI_{TS} is the average amplitude of the TSI variation over the most recent solar cycle, and ΔI_{TS} is the change in TSI between 1700, at the end of the Maunder minimum, and the most recent sunspot minimum. The same notation will be used in relation to other parameters later in this chapter, when I look at the ratio of the drift to recent cycle amplitudes $\Delta I_{TS}/ \delta I_{TS}$.

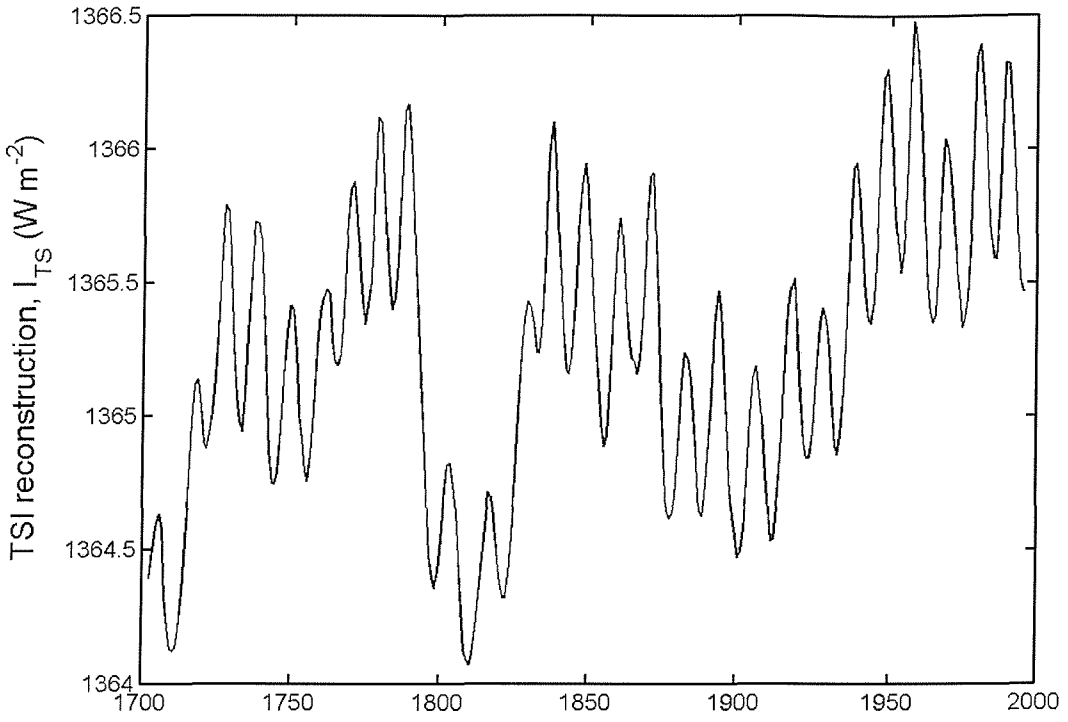


Figure 6.5. Reconstructed annual means of TSI calculated from the model of F_S variations by *Solanki et al.,* [2000].

Because they are fits to data from the past two solar cycles, the reconstructions of TSI shown in figure 6.4 have similar δI_{TS} of 1.1 Wm^{-2} and 1.2 Wm^{-2}. Both I_{Fs} and the *Lean* [2000a] reconstruction give minima in the Dalton solar minimum of near 1364.1Wm^{-2} in 1810 (compared with about 1365.6 Wm^{-2} in 1986 and 1996, at the last two solar cycle minima). However, the *Lean* variation gives somewhat lower values at the end of the Maunder minimum (1363.5 Wm^{-2} in 1710, as opposed to 1364.1 Wm^{-2} from annual I_{Fs}). Thus ΔI_{TS} is 2.1 Wm^{-2} for the Lean reconstruction and 1.5 Wm^{-2} for I_{Fs} (giving $\Delta I/\delta I$ of 1.91 and 1.33 respectively). In fact, the correlation gives slightly smaller long-term drift (smaller ΔI_{TS}) than all the other reconstructions that have been outlined in Chapter 1. It is interesting to note that the F_S intercept in figure 6.1, corresponding to no emerged magnetic flux, yields an absolute minimum of 1363.4 Wm^{-2}. Thus the I_{Fs} reconstruction suggests that even in the Maunder minimum there may have been some residual solar activity, corresponding to an associated TSI enhancement of 0.7Wm^{-2}. This is consistent with the ^{10}Be data sequence, which shows a solar cycle variation continuing during the Maunder minimum. This implies flux emergence continued during the Maunder minimum, even if none of that flux

was large and strong enough to show as sunspots. Even a relatively small magnetic enhancement, which may not create sunspots, could still cause faculae, and lead to an overall brightening. This has been investigated in further detail in section 6.4 and chapter 4. There is no way of quantifying the amount of flux threading the solar surface during the Maunder minimum period, but it would have been a relatively small amount by today's active solar standards. Several authors have set an upper limit to the drift in the surface field by assuming there was no magnetic activity, and thus no flux threading the surface, during the Maunder minimum.

Annual sunspot spread is also found to correlate very well with TSI (r = 0.87) and this comparison is shown in figure 6.6. As can be seen there is a linear best-fit comparison between the two parameters and this has been used to scale the annual sunspot spread values to give the variation in TSI.

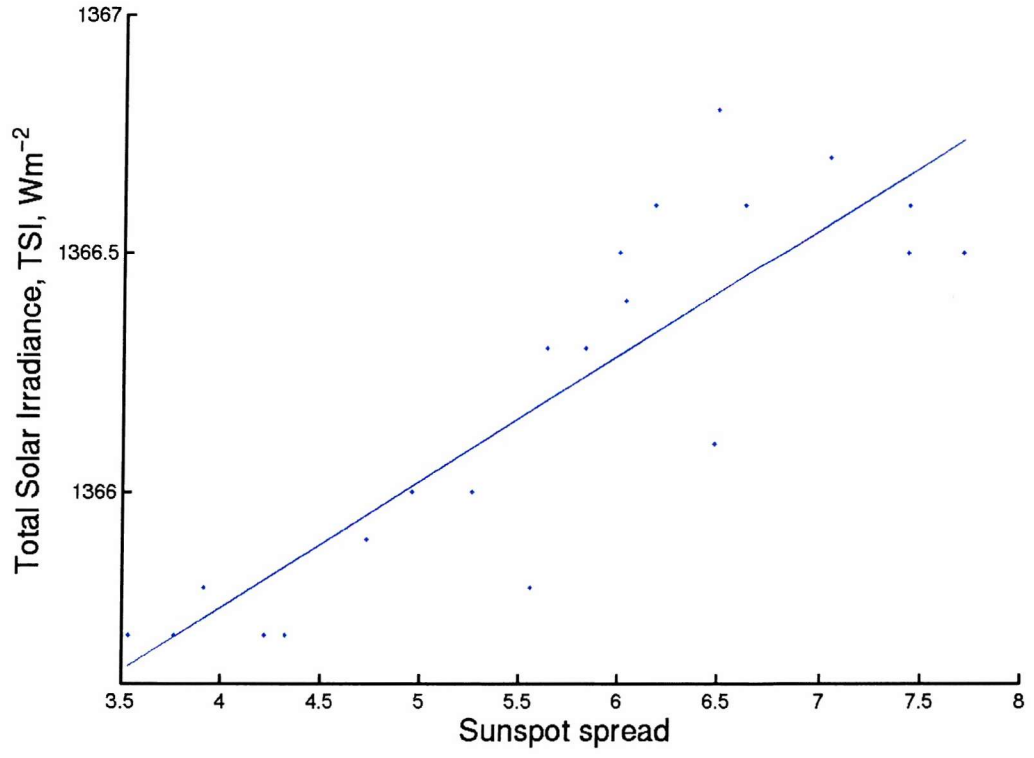


Figure 6.6. Scatter plot of annual means of TSI against sunspot spread.

Figure 6.7 shows annual variation of TSI for the period 1874-2000, and as can be seen it recreates the 11-year solar cycle variations, with the largest solar cycle amplitude (δI_{ss}) of 1 Wm^{-2} being found in solar cycle 19, which is a similar result to

the study by *Lockwood et al.* TSI is found to increase after 1874, with a drop around 1900, then a steady increase to its present day values, peaking in 1958. Taking the TSI value at the minimum in 1889 and the TSI value at the most recent minimum in 1996, I find a long-term drift of $0.3 Wm^{-2}$, which gives a long-term drift to solar cycle amplitude ratio of $\Delta I/\delta I = 0.3$.

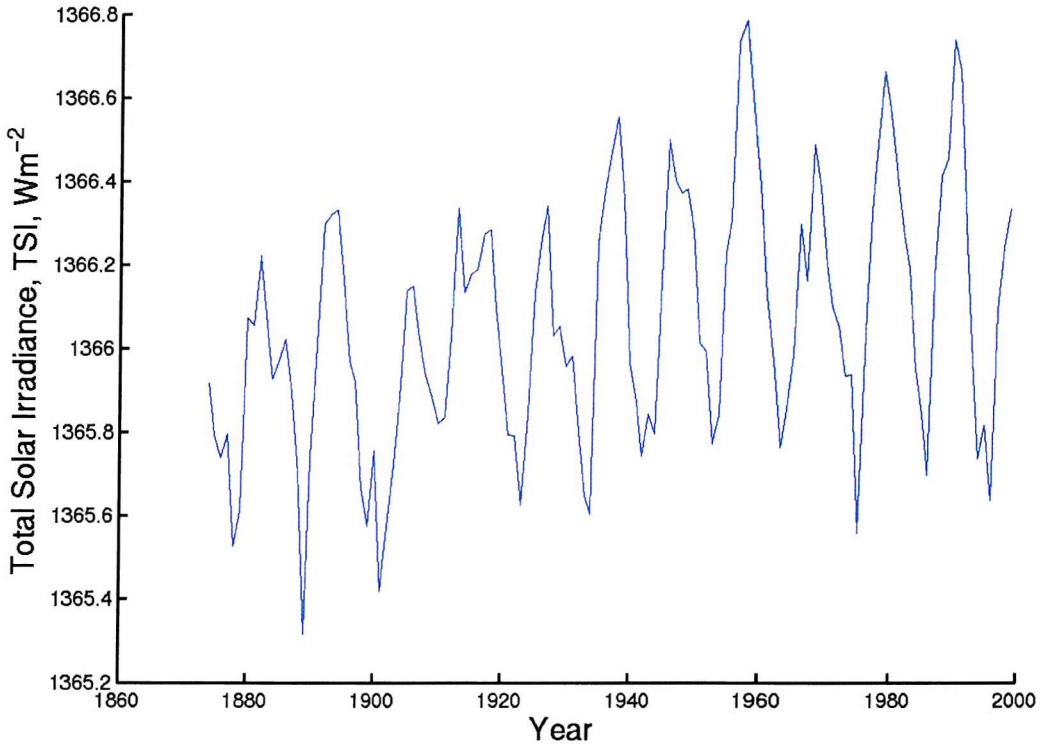


Figure 6. 7. Annual variations of TSI calculated from sunspot spread for the period 1874-2000.

Thus this reconstruction of TSI yields a lower long-term drift than other reconstructions. The reason for the lower long-term drift is the fact that at sunspot minimum there are relatively few sunspots and those sunspots that do occur all appear in a low narrow band at the solar equator. This leads to low values of spread, which do not increase in a proportional rate to other solar activity, such as F_s, although it increases more than for sunspot number, R and area A_S, which return to near zero at each solar minimum.

The key question is the same for the use of open flux as a proxy for TSI – specifically "what physical mechanisms could result in a dependence of TSI on the sunspot latitude spread?" As mentioned in Chapter 3, sunspot spread is thought to be

a good indicator of the total active region photospheric flux emergence rate, E_p, via both sunspots and faculae. Then when E_p is large, it will represent a large amount magnetic flux threading the active region solar surface, which will cause both more sunspots and more faculae. As will be outlined in sections 6.3 and 6.4, these active region sunspots and faculae are the main modulators of TSI over the solar cycle. As the number and size of sunspots increase then the photometric sunspot index (PSI) will increase causing a decrease in TSI, but at the same time faculae are increasing in numbers and area causing an increase in facular brightening (f_b) which will 'outweigh' the PSI, thus leading to an overall increase in TSI. Reliable measurements are only available for the past 20 years [*Chapman et al,* 1997] but it is thought that for every one sunspot there are approximately ten times as many equally sized faculae (see chapter 7). Thus although spread is calculated from sunspot latitudes and so defines the latitudes of the solar surface over which sunspots occur, it also defines a region over which active region faculae will also occur. Thus there is a plausible potential reason why sunspot spread maybe a better indicator of active region facular brightening than sunspots. The spread values show a long-term drift because of overlap between solar cycles when the first spots of a new cycle appear before the last spots of the old cycle have disappeared. Thus it may well be a good proxy for the extended solar cycle in smaller flux tubes, the overlap of which would give a long-term drift in facular brightening.

6. 2. 2. Galactic Cosmic Rays, Coronal source flux and Total solar irradiance.

Cosmic rays sweep out large, complex paths and so sample various regions of the heliosphere. Because the shield presented to cosmic rays, and which reduces the fluxes reaching Earth, depends on F_S it is to be expected, that there will be a strong anti-correlation between the coronal source flux and cosmic rays. The coronal source flux increases with sunspot number and this causes more cosmic rays to be deflected away from the inner solar system, meaning that fewer cosmic rays will be incident upon the Earth. Neutrons and muons, the products of galactic cosmic rays hitting Earths atmosphere have been recorded using neutron monitors since 1951 (see Chapter 2). *Lockwood* [2001] found a strong anti-correlation between the count rates

from Climax and Huancayo/Hawaii and the coronal source flux (r = 0.874 and 0.897, respectively).

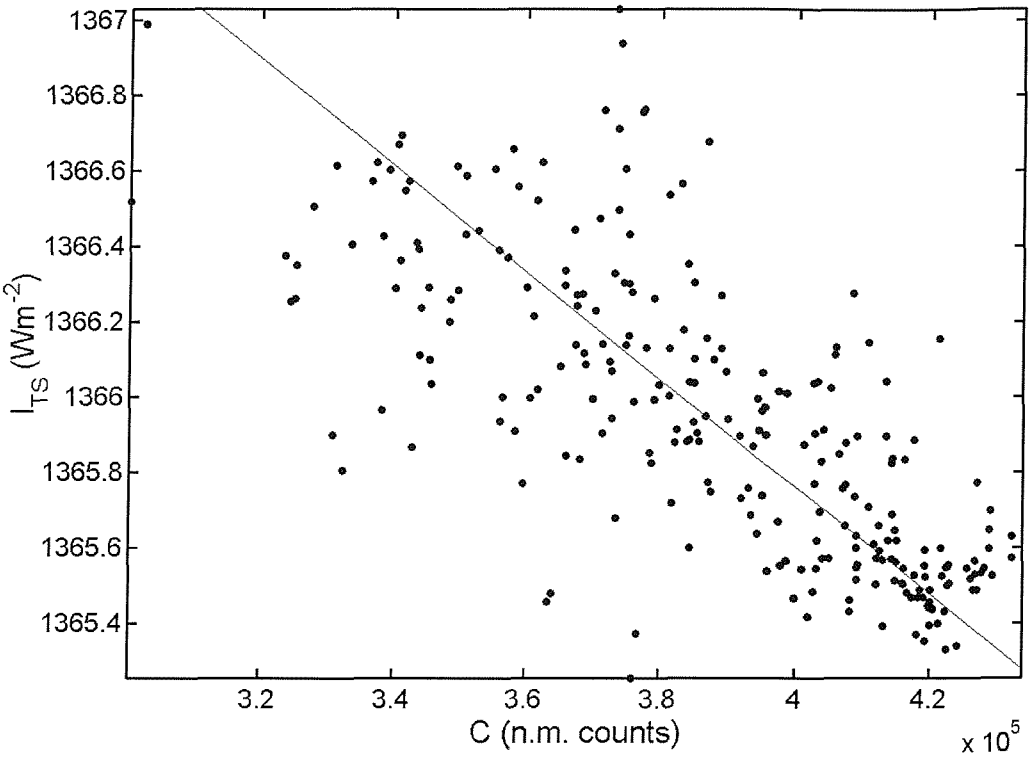


Figure 6.8. Scatter plot of monthly means of cosmic ray counts collected at the climax station and the monthly means of TSI.

Since such a high anti-correlation is evident between cosmic rays and the coronal source flux, it is clear there will also be an autocorrelation between fluxes and TSI. Cosmic rays were found to anti-correlate well with solar irradiance, with the best match being found between the Climax station and solar irradiance, with a lag of 3 months, with a correlation coefficient of r = 0.803. The scatter plot of high correlation can be seen in figure 6.8.

If this high anti-correlation between cosmic rays fluxes and TSI applies on century timescales as well as the decadal timescales of the data in figure 6.8, there should be a strong anti-correlation between TSI and the abundance of the ^{10}Be isotope present in ice cores, created by cosmic ray bombardment as discussed in Chapter 1.

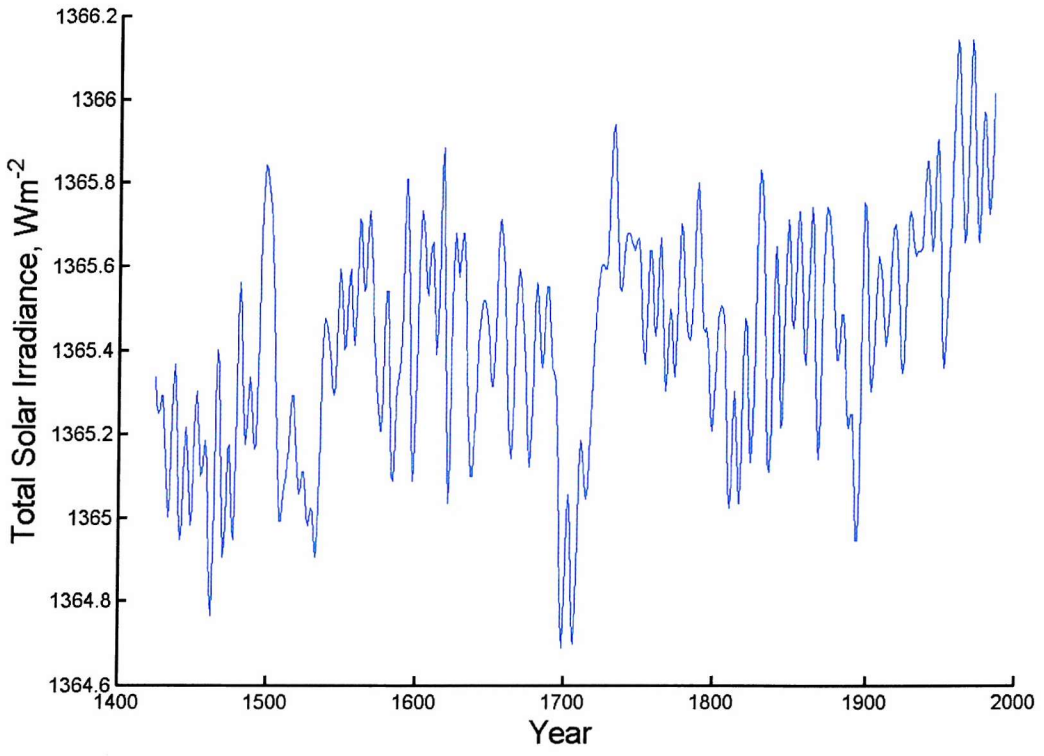


Figure 6.9. Annual variations of TSI calculated from variations in the ^{10}Be data.

This being the case, we could use this anticorrelation to calculate solar irradiance variations, from both the ^{10}Be and the F_s data. The Dye-3 Greenland ice core gives reliable ^{10}Be estimates back to 1424, and using the anticorrelation with TSI yields the variation shown in figure 6.9. This implies that TSI has increased since 1424, and a solar cycle variation is clearly evident, as is the long-term drift over the past 100 years that has been well documented by *Lean et al.* [1995] and *Lockwood et al.* [1999a]. The Maunder minimum period of low solar activity is clearly evident in figure 6.10 when the extrapolated TSI falls to 1364.7 Wm^{-2}. This value can be compared to other reconstructions, such as the Lean model, which puts a value on TSI at the end of the Maunder minimum of 1363.5 Wm^{-2}, and a value of 1364.1 Wm^{-2} from the reconstruction given in section 6.2.1. Although these values show a level of discrepancy, they all indicate a Maunder minimum solar irradiance that is significantly lower than present day values.

Cosmic rays are not just useful as solar proxies, they play important roles in the Earth's climatic system. Cosmic rays play a role in the global electric circuit, which is dominated by the occurrence of thunderstorms [*Bering et al.* 1998]. This is

because cosmic rays cause air ions to form in the subionic gap, which then allow current to flow between the tops of thunderclouds and the ionosphere and also between the ionosphere and the ground. It has also been postulated that global cloud is modulated by the cosmic ray flux [*Svensmark and Friis-Christensen*, 1997]. One problem with this correlation is that there is no clear distinction between different cloud types, which could be effected by cosmic rays. This was investigated by *Marsh and Svensmark* [2000], and they found that cosmic rays have a dominant effect upon (and thus correlate highest with) clouds that occur at low altitudes in maritime regions, and so is mainly associated with stratus and stratocumulus cloud types [*Svensmark,* 2000]. The mechanisms behind any effects on clouds by cosmic rays are not yet understood, although some explanations are coming forward. If cosmic rays are in fact found to play a role in the amount of global cloud cover, this would be an explanation of the links between climate indicators and cosmogenic isotopes revealed by several paleoclimate studies [*Bond et al.*, 2001]. However if this mechanism is not viable or not valid, different links between climate and cosmic rays must exist. For this reason I have been searching for the correlations between TSI and cosmogenic isotopes, cosmic rays and open flux described in this section and evaluating their implications for past TSI variations.

6. 3. Photometric Sunspot Index and Sunspot Darkening.

As discussed in Chapter4, the darkening of sunspots is produced by the interference with the upward heat flux by the strong magnetic fields, reducing the surface temperature by around 2000K. The Photometric Sunspot Index (PSI), P_{SI}, is a measure of the nett effect of the negative sunspot contrasts upon the Total Solar Irradiance (TSI) [*Hudson et al.* 1982]. Figure 2.2 shows the variation of PSI since 1981, as derived by the World Radiation Centre, Davos.

To reconstruct the variations of the PSI over the past 150 years, I have employed the composite of sunspot areas that have been made since 1874, as discussed in chapter5. Figure 6.10 is a scatter plot, which compares the mean monthly values of P_{SI} and sunspot area, A_s for the period 1981-2000 where the datasets overlap. The derivation of PSI, given in section 4.9.1 predicts that PSI will be

dominated by A_S, as shown in figure 6.10. The correlation coefficient is r = 0.990 (which has significance of 100% to within 6 decimal points). This excellent agreement allows us to continue the P_{SI} sequence back in time, using A_S as a proxy. The solid line in figure 6.10 shows the best-fit linear regression between P_{SI} and A_S, and this fit is used to calculate the variation of PSI from the observations of sunspot areas. The results for 1874-2000 are shown in figure 6.11. As well as the well-defined 11-year solar cycle, the PSI at sunspot maximum is found to have increased considerably over the past 130 years with the highest being found in solar cycle 19, with more than double the maximum darkening found in cycles 12 and 13. Because the sun is almost free of all spots at sunspot minimum, P_{SI} returns to close to zero at each solar minimum. However, careful inspection of figure 6.11 shows that minima in P_{SI} and A_S have become marginally larger and shorter over the past 100 years.

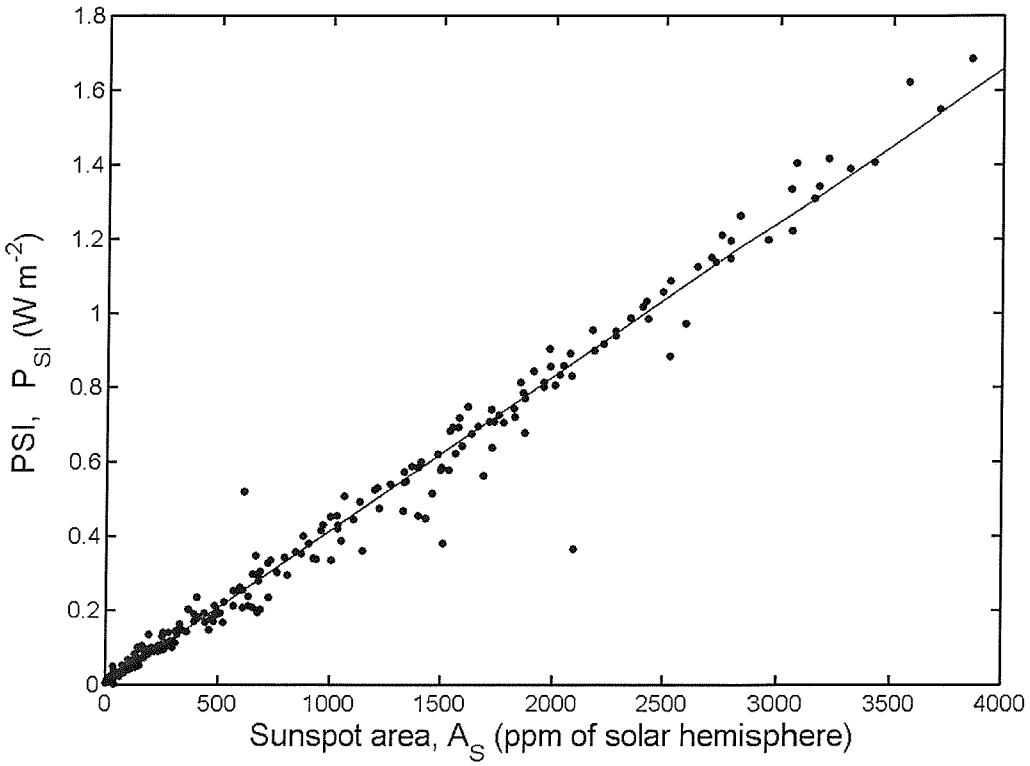


Figure 6.10. Scatter plot of monthly means of the photometric sunspot index, P_{SI}, against sunspot area A_s for the period 1981-2000. The solid-line is the best fit-linear regression.

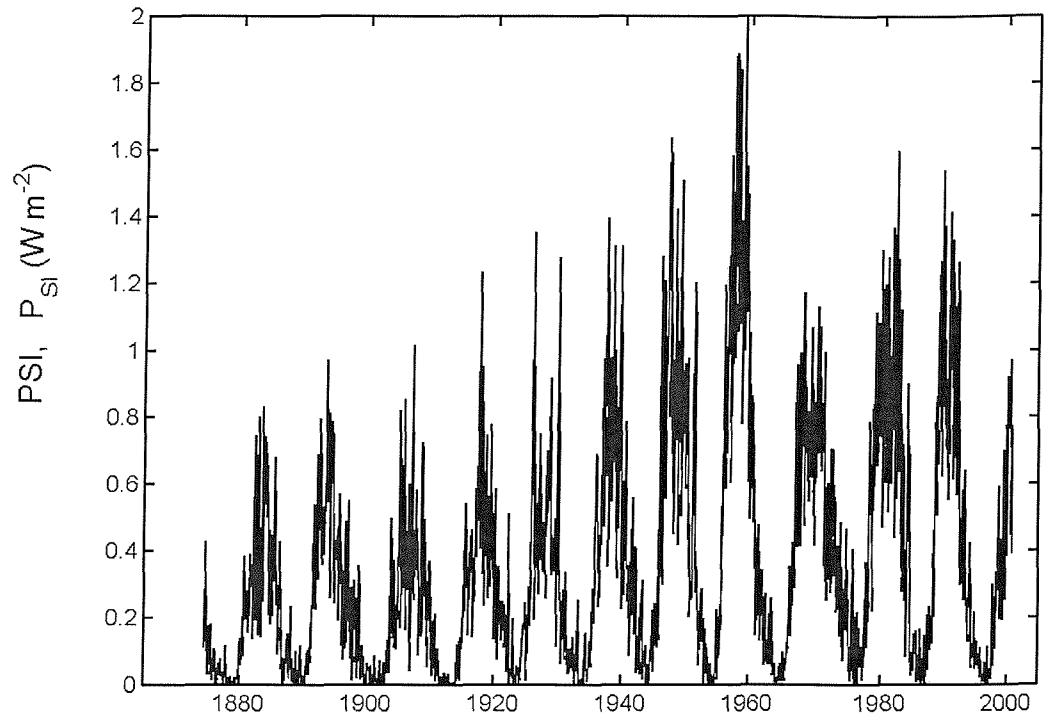


Figure 6.11. Monthly variations of the photometric sunspot index P_{SI}, calculated from the monthly means of observed sunspot areas for the period 1874-2000, using the composite sunspot group dataset.

Unfortunately, sunspot area measurements only extend back to 1874, and so to continue our data set of PSI further back in time I used the only available proxy, the sunspot number, R. Figure 6.12 shows the scatter plot of monthly Sunspot number, R, against P_{SI} for the period 1981-2000. As can be seen the agreement is not quite as good as between P_{SI} and A_S, the overlapping R and P_{SI} data sets giving a correlation of r =0.918. Figure 6.12 also shows the variation is not quite linear and the regression line shown is a cubic least-squared fit, which has been used so that the behaviour at lower values is better reproduced. The best fit of the two data sets is used to create a data set of the monthly variations in PSI dating back to the start of monthly R values in 1748. This can be seen as the red line in figure 6.13.

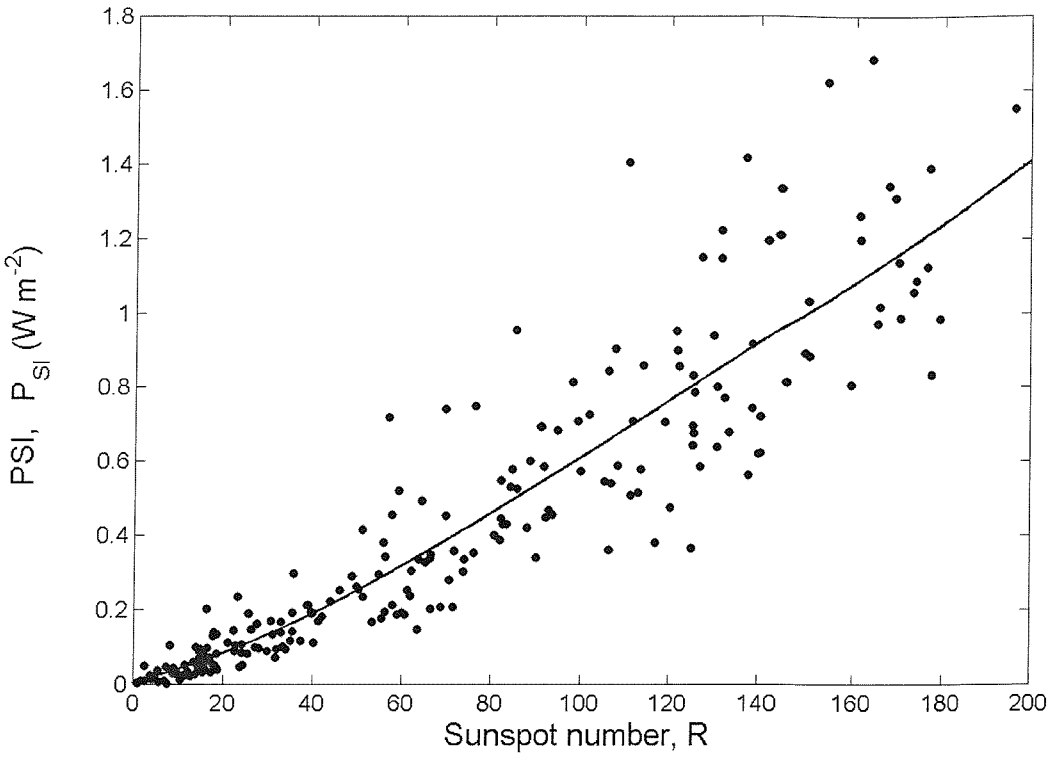


Figure 6.12. Scatter plot of the monthly means of sunspot number R against the photometric sunspot number, P_{SI}, for the period 1981-2000. The solid line is the best-fit cubic regression.

As a test of the use of R as a proxy for PSI, in figure 6.14 I compare the monthly values for 1874-2000 derived from the sunspot area, A_S (using the linear regression shown in figure 6.10) with those derived from the sunspot number, R (using the cubic regression shown in figure 6.12). The best-fit line would ideally have slope unity, but in fact has a slope 11% greater than this. The PSI value derived from R is corrected by this factor, so that the two data sets are consistent where they overlap.

Annual values of sunspot numbers are also available from 1700 onward, and these are employed to create an annual data set of PSI, in a similar way to the monthly values outlined above. These annual values are shown as the black line in figure 6.13. The agreement with the monthly values is good. From this reconstruction it can be seen the δP_{SI} is 1.0 Wm^{-2} but the long-term drift in solar minimum values is very small, $\Delta P_{SI} = 0.02 \mathrm{Wm}^{-2}$ (giving $\Delta P_{SI} / \delta P_{SI} = 0.02$).

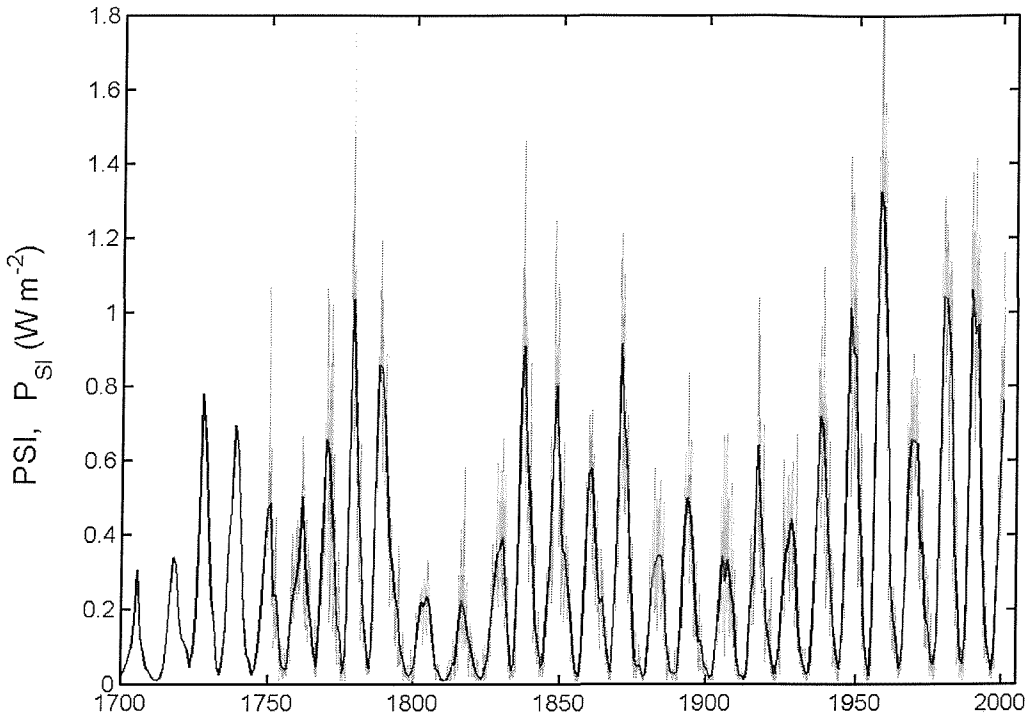


Figure 6.13. Variation of PSI since 1700. The grey line is the calculated using monthly means of sunspot number, *R*, black line is calculated from annual means of *R*.

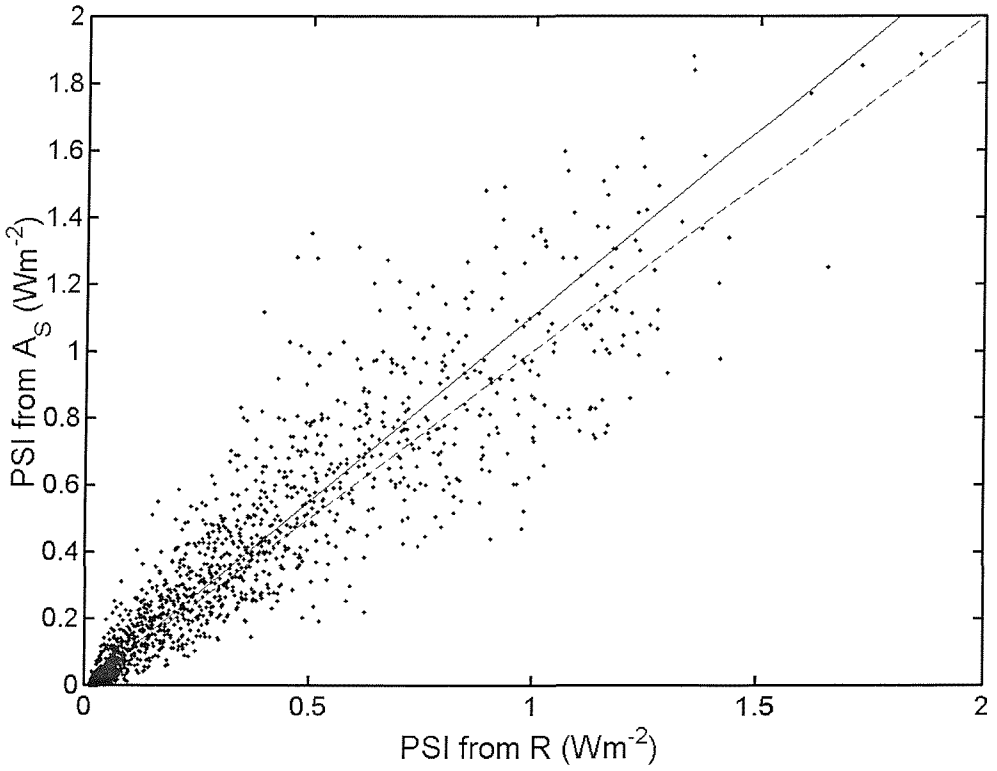


Figure 6.14. Comparison of monthly P_{SI}, calculated from sunspot area and *R*.

6. 4. Facular Brightening

Having reconstructed PSI and TSI, on both annual and monthly timescales for as far into the past as possible, from sunspot records, the next task was to obtain a model of brightening caused by the effects of active region faculae, the network and any other mechanisms relevant to the quiet Sun. The role of faculae in TSI variations was first noted by *Chapman* [1987]. Given that there is good evidence that PSI quantifies the net darkening effect of sunspots, this section can be viewed as a quantification of the brightening effect required by TSI reconstructions [*Frölich and Lean,* 1998a; b; *Fligge et al.,* 1998]. A simple equation gives TSI as:

$$I_{TS} = Q + f_b - P_{SI} \tag{6.1}$$

where f_b is the net brightening effect due to faculae and Q is the brightness of the "quiet" Sun, I_{TS} is the total solar irradiance, and P_{SI} is the Photometric sunspot index. The definition of Q used, means that f_b includes the effects of faculae in both active regions and in the network [*Solanki and Stefano,* 1984; *Solanki and Brigljević*, 1992; *Ortiz et al.,* 2002; *Wenzler et al.,* 2002] Q would also include any other magnetic effects, for example "shadow" effects in effective photospheric temperature due to magnetic field changes in the convection zone. Effects of unresolved and weak field structures in the intra-network photosphere and "ephemeral" regions are also included in Q.

Because Q is unknown, I initially group f_b and Q into a single variable which, from equation (6.1) I can obtain by adding the PSI to the observed TSI. Figure 6.15 shows monthly average ($Q + f_b$) values calculated from the composite satellite measurements of TSI, and the *Frohlich* PSI data series (see figure 2.2), for the period 1982-2000.

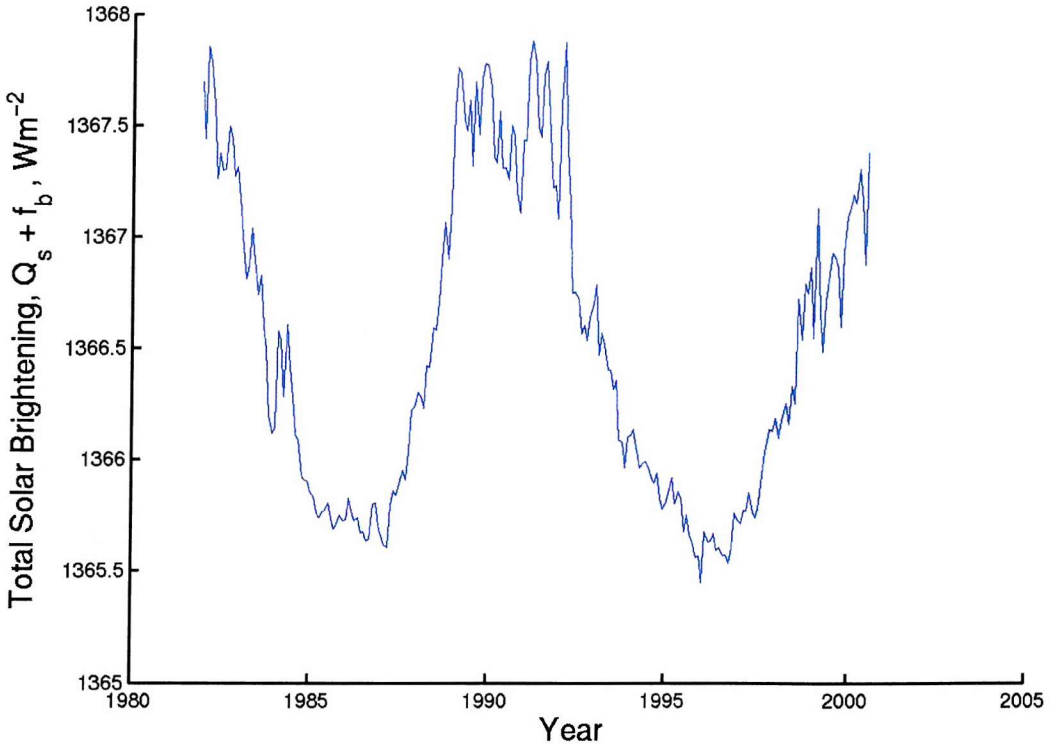


Figure 6.15. Variation of the monthly means of quiet sun plus facular brightening for the period 1981-2000.

Measured values of TSI and PSI only go back to 1978 and 1981 respectively, and so to continue the data set of $(Q + f_b)$ further back in time the data sets of TSI and PSI calculated in the previous sections were used as inputs into equation (6.1). Figure 6.16 shows calculated monthly values of $(Q + f_b)$ for the period 1874-2000 in grey. The magnitude of the factor $(Q + f_b)$ is dominated by the quiet sun irradiance, Q, but how much of its variation over recent cycles is introduced by f_b [*Solanki and Fligge*, 2002] and how much is introduced by Q [*Kuhn et al.*, 1999] is a key point of discussion. If the correlation between I_{TS} and F_s is valid on all timescales, the strong implication would be that all of the changes in $(Q + f_b)$ are magnetic in origin (Q varies because of the effect of convection zone fields and/or the effect of small intra-network magnetic fields in the photosphere f_b varies because of facular flux tubes in active regions and the network), implying that shadow and surface magnetic effects are the only possible factors.

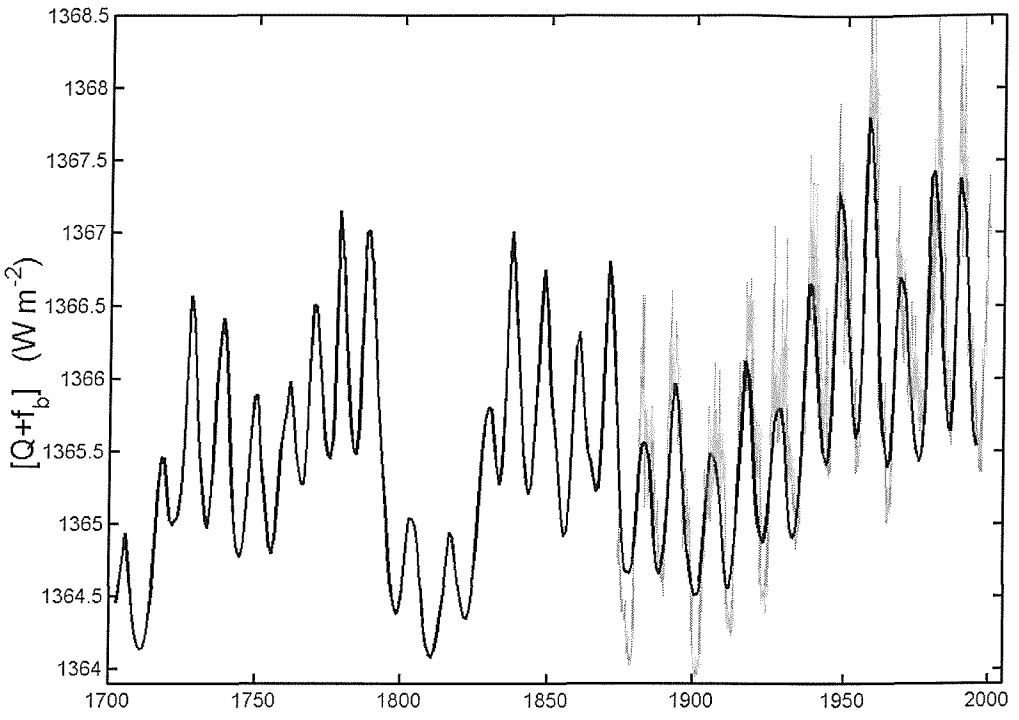


Figure 6.16. Variation in the quiet sun TSI plus facular brightening, (Q+f_b). In grey are monthly values for the period 1874-2000; in black are annual values for 1700-2000.

As can be seen ($Q + f_b$) has risen steadily since 1900, with the minimum value in the most recent cycle being 1.5 Wm^{-2} larger than the minimum point in 1900 $\Delta(Q + f_b) = 1.5$ Wm^{-2}. The solar cycle amplitude of ($Q + f_b$) has also increased since 1900 with values in the most recent cycle being larger than those in the 1900 solar cycle, by around $1Wm^{-2}$. Figure 6.17 compares the values of ($Q + f_b$) calculated from the measured values of TSI and PSI, with the values of ($Q + f_b$) calculated from the proxies of PSI and TSI for the period 1981-2000. The two data sets show good agreement, with a correlation coefficient of r = 0.9261, and the line of best fit between the two data sets shows that although there is some scatter, the two data sets match up extremely well.

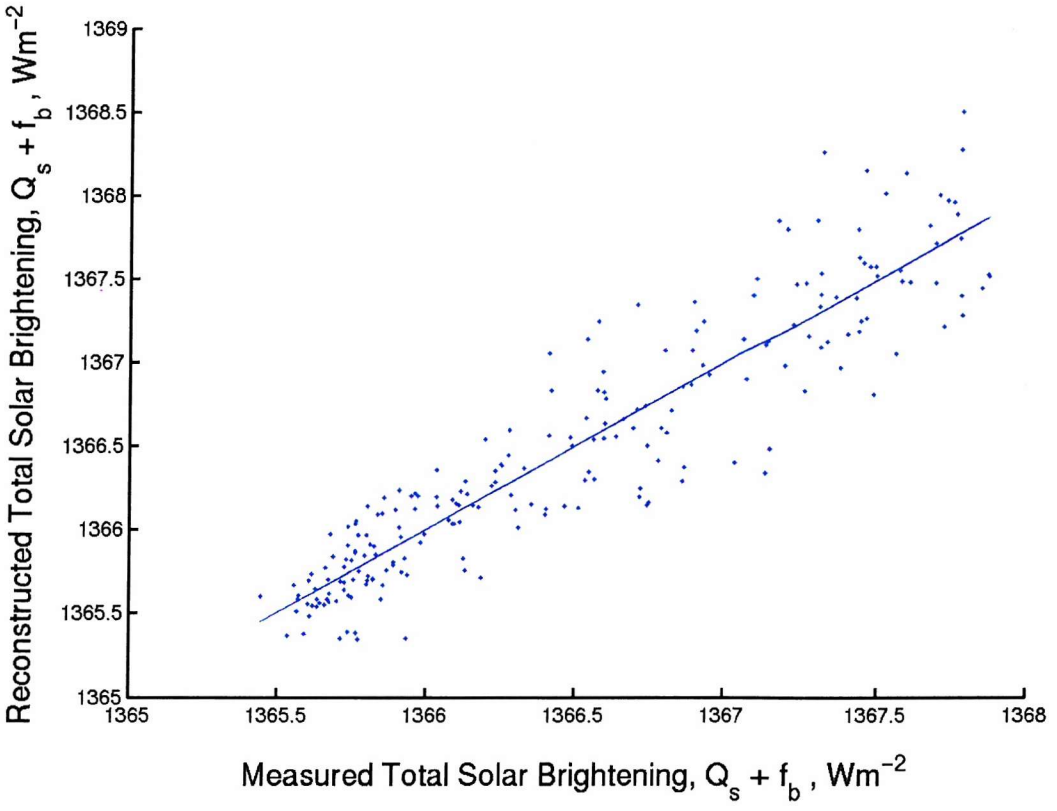


Figure 6.17. Scatter plot of the monthly values of facular brightening plus quiet sun (Q+f_b) calculated from observed and reconstructed values.

Chapman et al. [1997] have made measurements of facular area at the San Fernando solar observatory, and it has been shown that facular area A_f measurements are an excellent proxy for facular brightening in active regions f_{bA}. Figure 6.18 shows monthly values of facular area in active regions A_f, plotted against the monthly values of ($Q + f_b$) derived from the proxies of TSI and PSI for the period 1978-2000. The two data sets show excellent agreement, with a correlation coefficient of r = 0.97. I define f_b to be the sum of magnetic brightening outside active regions in the network, f_{bN} and by active region faculae, f_{bA}. The correlation shown in figure 6.18 is slightly lower than would be expected for f_{bA}, due to the presence of magnetic effects outside active regions, for example the brightening effect of faculae in the network f_{bN}, bright rings around sunspots and any solar cycle variation in Q. Figure 6.18 shows a slight indication of a change in slope at the very lowest values, consistent with $A_f = 0$ intercept being equal to the minimum observed ($Q + f_b$) of 1365.6 Wm^{-2}; for which limit $f_b = f_{bN}$, and $f_{bA} = 0$; i.e. any brightening over the quiet-sun level Q must all be supplied by network faculae. Figure 6.18 gives ($Q + f_b$) = 1365.6 Wm^{-2} for recent solar minimum. The largest data points in figure 6.18 are at 1367.9 Wm^{-2} showing

that the peak brightening over the minimum value is $\delta(Q + f_b)$ =2.3 Wm^{-2}. *Foukal et al* [1991] noted that not all of this $\delta(Q + f_b)$ could be attributed to active-regions faculae and found that most of the remainder was associated with network and δf_{bN}; on the other hand, *Kuhn and Liberecht* [1991] argue that there is a solar cycle variation in Q.

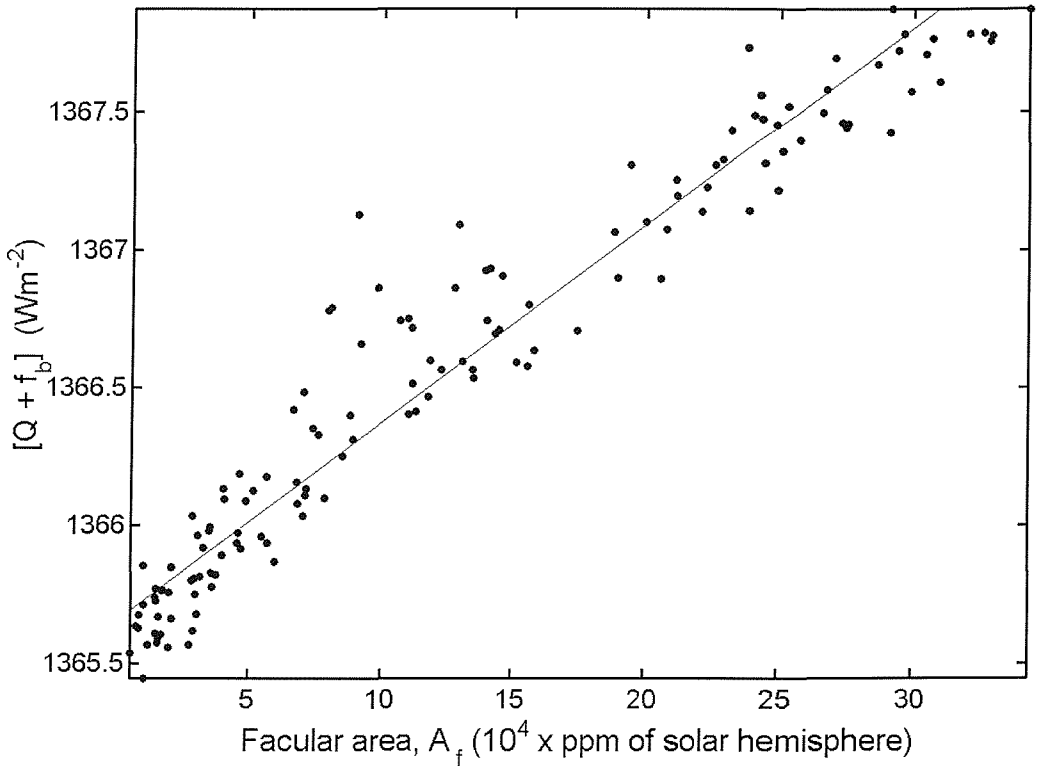


Figure 6.18. Scatter plot of monthly values of facular area, A_f, against facular brightening plus quiet sun TSI, $(Q+f_b)$ for the period 1874-2000.

It is also possible to create annual values of $(Q + f_b)$ from our annual values of PSI and TSI, using equation (6.1), for the period 1702-2000. These are plotted in black in figure 6.16, and are found to vary in a similar fashion to the monthly values for the overlapping period. As discussed previously, there is evidence that there was still cyclic magnetic activity in the Sun during the Maunder minimum. Because there were almost no visible sunspots and active regions during the Maunder minimum, it is reasonable to assume that $f_{bA} = 0$ in the Maunder minimum and this variation must be due to the network flux tubes, intra-network flux or ephemeral flux. Thus they could be associated with a corresponding f_{bN} and/or Q variation. So I cannot assume that, even during the Maunder minimum, f_b fell to zero. However if I take f_{bN} to be zero at

the times of minimum reconstructed I_{TS} during the Maunder minimum, I am able to set a maximum value on Q of 1364.1 Wm^{-2}. Figure 6.16 shows that this minimum I_{TS}, fully quiet sun would have existed only during the Maunder minimum and also briefly during the Dalton minimum. From this I derive a value of $\Delta(Q + f_b) = 1.5 Wm^{-2}$. The ratio of the long-term to solar cycle brightening is $\Delta(Q + f_b)/\delta(Q + f_b) = 0.65$. The ratio is small because of the dominant effect of active region faculae on the solar cycle.

6. 5 Potential Surface magnetism Effects on the Quiet Sun

Harvey [1994] analysed Kitt Peak magnetograms and found that the average photospheric magnetic field is higher at sunspot maximum than at minimum by a factor of about 3. A radial field strength of $|B/\mu|$ gives an (unsigned) field seen by a magnetogram of $|B|$ and averaging the observed component over all pixels covering the disk yields $|B|_{av}$, which varies between 7G and 23G. For these, a complete absence of photospheric flux in the Maunder minimum yields the limit $\Delta|B| \leq 7G$ and hence $\Delta|B|/\delta|B| \leq 0.44$. This is considerably smaller than $\Delta I_{TS}/\delta I_{TS} = \Delta F_s/\delta F_s = 1.33$ required by the TSI reconstruction and somewhat smaller than the $\Delta(Q+f_b)/\delta(Q+f_b)$ of 0.65 derived in the previous section.

However, within the 1"×1" resolution of the Kitt Peak magnetograms, regions of opposite field will cancel each other out and give smaller $|B|$ than would be observed with higher spatial resolutions; the importance of this effect being higher in lower resolution magnetograms. *Kivrova et al.* [2002] have used high-resolution SoHO MDI magnetograms (0.045"x0.045" resolution) to show that, pixels that are smaller in area by a factor 500 than for the Kitt Peak data give $|B|$ that is larger by a factor of 2.5 for the quiet sun. In comparison, the equivalent factor for the larger-scale magnetic fields of active regions is only 1.1. The quiet-Sun $|B|$ only increases slightly over the solar cycle and thus global magnetograms of the same resolution as the (localised) high-resolution MDI data would have given $|B| \approx 2.5 \times 7 = 17.5G$ at recent solar minima whereas the solar maximum values would have been of order 17.5x1.1(23-7) = 35G. Thus $\delta|B| \approx 17.5G$, $\Delta|B| \leq 17.5G$ and $\Delta|B|/\delta|B| \leq 1$. This limit is closer to the ratio $\Delta F_s/\delta F_s = 1.33$ for the observed open flux variation and could be

closer still for yet higher resolution magnetograms. Some of the recent modelling work predicts that the open flux variation will indeed be similar to that of the total surface flux ($\Delta|B|/\delta|B| = \Delta F_s/\delta F_s$ $\Delta F_s/\delta F_s$). [*Solanki et al.* 2002]. For a spatially uniform field, the average radial field $|B/\mu|_{av}$ is 1.5 $|B|_{av}$ and the ratio of the (signed) open flux to the total (signed) photospheric flux is $F_s/(2\pi R_s^2 x|B/\mu|_{av})$. With the estimated values given above it was found that this fraction is 5.0% at solar minimum and 4.3% at solar maximum. Thus, with yet higher resolution images, it is plausible that the surface flux could be proportional to F_s such that it's long-term variation could indeed be very similar to the open flux variation shown in figure 6.1, with a long-term rise in the quiet sun average field. There is as yet, no observational evidence that this is indeed the case, however the possibility should be investigated as a potential physical basis for the open flux - TSI correlation.

The additional surface flux revealed by Kivrova et al. is in the form of small-scale magnetic structures of the quiet sun. these have been linked to bright intra-network elements, seen in Ca II K chromospheric proxy emissions. However, the photospheric variations associated with these quiet Ca II K variations are 50 times higher than for the same Ca II K change in active regions [*Kuhn et al.,* 1999]. Thus magnetic flux tubes on these smaller scales are a possible cause of the reported non-facular quiet sun variations [*Kuhn and Libbrecht,* 1991].

6.6 Discussion

The successful modelling by *Fligge* et al. [1998] leads to the implication that TSI variations over the past two solar cycles are due to photospheric magnetic flux tube effects alone [*Solanki and Fligge,* 2002; *Kivrova et al.,* 2003]. If there are no changes in the quiet Sun, the maximum Q of 1364.1 Wm^{-2} derived here would apply at all times since the Maunder minimum. The implications of this are investigated, before considering likely variations in Q.

6.6.1 Implications of constant Q

From the limiting value when the active-region facular area goes to zero it can be deduced that $(Q+f_b)$ in recent cycles was equal to 1365.6 Wm^{-2}. If the quiet Sun contribution is constant, a magnetic brightening of network faculae f_{bN}, of at least 1.5 Wm^{-2} at recent solar minima, such that Δf_{bN} =1.5 Wm^{-2}, is required.

Recently, *Walton et al.* [2003] have employed a database of individual solar features, which has been compiled from the full disk photometric images taken at the San Fernando Observatory (SFO) since 1989. This data set has been used to model the fraction of variation of TSI that is due to solar features of various sizes. The data shows that large-scale solar features dominate the change in f_b between solar maximum and solar minimum and f_{bN} produces about 15% to 25% of the total TSI changes over the solar cycle, consistent with the analysis of *Foukal et al.* [1991]. If an average of the two limits (i.e. $\delta f_{bN}/\delta I_{TS}$ =0.2) is used, with the observed δI_{TS} of 1.2 Wm^{-2} I find that δf_{bN} = 0.24 Wm^{-2} and, because $\delta(Q+f_b)$ = 2.3 Wm^{-2}, δf_{bA} = 2.06 Wm^{-2} for δQ = 0. This can be compared with the required Δf_{bN} = 1.5 Wm^{-2} for ΔQ = 0. By comparison, the open flux proxy yields $\delta F_s = 1.73 \times 10^{14}$ Wb and $\Delta F_s = 2.59 \times 10^{14}$ Wb. Thus $\Delta f_{bN}/\delta f_{bN} \geq 6.2$, considerably greater than the corresponding ratio $\Delta F_s/\delta F_s$ = $\Delta I_{TS}/\delta I_{TS}$ = 1.5, i.e., the reconstruction with constant Q calls for a disproportionately large long-term change in network faculae Δf_{bN}.

The question arises that if there is sufficient brightness in the network faculae seen at recent solar minima, such that should it be removed the irradiance falls to, or below, the values predicted by the reconstruction for the Maunder minimum (i.e., at recent solar minima f_{bN} must exceed the required Δf_{bN} = 1.5 Wm^{-2}). For example, analysis of Ca-II K plage emissions from the quiet network chromosphere, as observed in 1905 and 1984 suggests insufficient change to explain the required Δf_{bN} [*Foukal and Milano,* 2001]. However, this raises secondary questions about the validity of these chromospheric plage emissions as a proxy of the modulation of emissions from the underlying photosphere. There are also considerations here of the changes in image sensitivity and resolution as instruments has developed.

However, a similar conclusion was reached in the study discussed above by *Walton et al.* [2003]. They estimated that the reduction in f_{bN} if all photospheric magnetic features are removed would only be of order $\Delta f_{bN} \approx 0.4$ Wm^{-2}, considerably less than the Δf_{bN} of 1.5 Wm^{-2} required by the TSI reconstructions discussed here. These considerations indicate an inconsistency and lead to the implication that changes in network faculae are insufficient to explain the long-term drift and that it is necessary to also invoke changes in Q.

6.6.2 Implications of variations in Q

The upper Δf_{bN} limit of 0.4 Wm^{-2}, discussed above, would require $\Delta Q \geq 1.1$ Wm^{-2}. Because the other limit is $\Delta f_{bN} = 0$, I have a requirement for a long-term quiet-sun irradiance change $\Delta Q = 1.3 \pm 0.2$ Wm^{-2}. In this section the implications of such a change are looked at, by investigating the dependence of irradiance contributions on surface magnetic fields that this implies, using the hypothetical high-resolution magnetogram results discussed in section 6.5. A linear dependence is assumed of brightness and magnetic field changes, both for faculae $\chi_f = df_b/d|B|_{av}$ (assumed to apply for both network and active region faculae) and for the quiet sun $\chi_Q = dQ/d|B|_{av}$. Considering recent solar cycles I get $\chi_f = \delta f_{bA}/\delta|B|_{av} = 2.3/17.5 = 0.131$ Wm^{-2}G^{-1}. Because $\Delta|B|_{av} \leq 17.5$ G, and $\Delta Q = 1.3 \pm 0.2$ Wm^{-2}, it is found that $\chi_Q = \Delta Q/\Delta|B_Q| \geq 0.074 \pm 0.011$ Wm^{-2}G^{-1}. This limiting value for χ_Q is roughly half that for the faculae, χ_f and applies if there is no residual quiet-sun flux at the Maunder minimum. A larger χ_Q of 0.138 Wm^{-2}G^{-1} (similar to that for faculae) would leave a residual $|B|$ of 5.1 G at the Maunder minimum, the removal of which would cause Q to fall further to the extrapolated limit of 13634 Wm^{-2} corresponding to zero open flux, $F_s = 0$. If these χ_Q estimates are reasonable, a scenario explaining the correlation between F_s and TSI may be possible, through a relation of F_s to the surface field and the effect of surface field on both facular and quiet-sun emissions.

Note however, the complex mixture of effects needed to give the I_{TS}-F_s correlation, with active region changes and high surface fields dominating recent solar cycle variations and the quiet sun and small-scale structures dominating the long-term drift. This means the overall linear variation can only be approximate and a non-linear

variation is likely to be more appropriate (depending on if χ_f and χ_Q are constant, and on their relative values).

6.6.2 Implications of variations in Q.

A strong and significant correlation is found over recent solar cycles between the open solar flux and total solar irradiance. Although it is not yet understood, simple extrapolation based on this correlation provides a TSI reconstruction that is similar to others, based on a variety of other proxies. If these other reconstructions are valid, the correlation between TSI and open flux must apply on century as well as decadal timescales. Given that open flux is closely anticorrelated with the flux of cosmic rays reaching Earth, this also explains the strong anticorrelation observed between TSI reconstructions and cosmogenic isotopes on both decadal and century length timescales and provides justification for the many paleoclimatic studies that make use of cosmogenic isotopes as indicators of TSI.

The photometric sunspot index has been found to correlate extremely well with sunspot area and a PSI data series has been constructed based on this correlation, for the period 1874-2000. This reconstruction shows that PSI has steadily increased over the solar cycle since 1900, in a similar fashion to the rise in TSI, with the maximum amount of P_{SI} being found in solar cycle 19. This reconstruction was continued in monthly form back to 1748 and in annual form to 1700, using sunspot number, which has shown a strong correlation to measured P_{SI}. These datasets are remarkably useful, as now the darkening effect of sunspots upon the TSI has been accurately quantified.

Using the reconstructions of TSI and PSI, an estimated facular brightening and quiet sun data series ($Q + f_b$), was developed on both monthly and annual timescales, which was found to correlate extremely well with known indicators of facular brightening such as facular area. This reconstruction of ($Q + f_b$) was found to have a long-term drift similar to that found in TSI, This upward drift in TSI and ($Q + f_b$) was analysed, and it was concluded that the long-term drift in TSI and ($Q + f_b$) is too great

to be attributed to network faculae alone, and that the quiet sun Q_s must play a dominant role in the long-term increases found.

Chapter 7: Irradiance reconstructions using contrasts of magnetic features from SoHO data.

7. 1. Chapter overview.

In Chapter 4, I reviewed the darkening effect of large magnetic flux tubes (sunspots) and in chapter 6 I showed how we could use various historic data, to estimate past values of the photometric sunspot index, provided the various datasets were intercalibrated with care (see chapter 5). In order to reconstruct past variations of the total solar irradiance, we also need to evaluate the brightening effect of small flux tubes (faculae). In this section, I evaluate the contribution of faculae associated with active region sunspot groups and explore ways that we can estimate the facular brightening over the past 150 years.

The equations of *Ortiz et al.* [2002] enable the calculation of the contrast C_{MDI} for any one pixel of a full-disc MDI image in the continuum emission around 676.8 nm due to small flux tubes which contribute to facular brightening. Figure 7.8 in section 7.3 shows how C_{MDI} varies with the field in the magnetogram signal, B_{MDI} and the position on the solar disc, μ (and hence the radial field value, B_{MDI}/μ). The contrasts are relative to a field-free quiet sun intensity, corrected for limb darkening using the function derived by *Neckel and Labs* [1994], as discussed in Chapter 4. At the wavelength of 676.8nm facular contrast is close to its mean volume, averaged over the UV/visible/IR spectrum [*Unruh et al.,* 1999]. Thus, like *Ortiz et al.* [2003], I here use C_{MDI} values for TSI calculations. In general a small correction $C_{all\ \lambda}/C_{MDI}$, which may be a weak function of μ is required, but this is assumed to be unity in this section.

In order to exploit the *Ortiz et al.* contrast, we need to know the distribution of B_{MDI}/μ values in active regions and in the quiet Sun. This will enable us to evaluate the distribution of pixel contrasts in active regions.

7. 2. Distributions of magnetic fields in small flux tubes

Studies have been conducted by *Ortiz et al.* [2000, 2003] on one specific active region (AR NOAA 7978) using whole-disc MDI images and quantified the field distribution within it. This active region crossed the solar disc during the 1996 solar activity minimum and was the only one present on an otherwise relatively featureless Sun. Thus *Ortiz et al.* were able to study the effect of the facular brightening associated with this region on the observed total solar irradiance. *Harvey and Hudson* (2000) noted that a complex of active regions first appeared in Carrington Rotation (CR) 1908. During CR 1911, on 4 July 1996, a strong new centre of magnetic activity appeared within this complex to the east of the disk centre before rotating onto the far side of the Sun. This AR was catalogued as AR NOAA 7978. It can be seen in the daily total sunspot group areas ($\Sigma_i A_{GI}$) plotted in blue in figure 7.1. It contributed all the total group area when is rotated back onto the visible disk. The sunspot group, as seen during the next two rotations, is shown in figure 7.2. By 3 September, the spots have almost all disappeared, leaving faculae in the area where the active region had been: we here loosely refer to this as ephemeral flux. SoHO MDI magnetograms, observed when active region AR NOAA 7978 was near the centre of the disk during Carrington rotations 1911-1916, are shown in figure 7.3. It can be seen that the flux tubes of the AR become smaller and more dispersed as it evolves and spreads out and the large flux tubes (sunspots) disappear..

In the right hand panels of figure 7.1, the times when the centre of AR NOAA 7978 were on the farside of the sun are marked in grey and the times when it was close to the centre of the visible disk (about which the distributions shown to the left of the figure were observed) are given by the dashed lines. The evolution of the sunspot group described above can be seen, with almost no sunspots being observed after they fade during CR1913. The red plot to the right of figure 7.1 shows the sum of the TSI and PSI, which by equation (6.2) equals the sum of the quiet sun irradiance and the total facular brightening ($Q+f_b$). The variations with solar rotation show how this active region is dominated both the sunspot group data and the irradiance variations at this time, with the irradiance elevated only when AR NOAA 7978 was on the visible disk. Initially, the facular enhancement is greatest when the active

region is near the limb, giving the characteristic double-peaked temporal variation during CR1912. This is also present in CR1913, but is less marked as the region expands in area. By CR1914, the remnant of the AR has expanded to the extent that the μ-dependence of the excess emission of the individual flux tubes is smeared out by the variety of μ values present. Note that the excess facular emission persists after the sunspots have faded, but was also present before the main sunspots appeared during CR1911. As the region faded the distributions of radial field $N_{MDI}(B_{MDI}/\mu)/N$ evolve back towards the quiet sun form: the quiet sun distribution for this time, $n_{QS}(B_{MDI}/\mu)$ / N, as given by *Ortiz et al.*[2003], is shown in blue. (N is the number of MDI pixels showing a radial field B_{MDI}/μ values and N is the total number of MDI pixels in the region in question).

The left hand panel of fig 7.1 shows radial distributions of magnetic fields observed near disk centre. Magnetogram signal B_{MDI} is converted into radial field (B_{MDI}/μ) by assuming that the field is radial. Observed distributions are shown as black histograms for Carrington rotations 1911-1916 (July-December, 1996). The numbers of pixels are here for 10G bins (B_{MDI}/μ). The red and green distributions are polynomial fits to the outward and inward halves (respectively, $0 < B_{MDI} < 1000$G and $-500\text{G} < B_{MDI} < 0$) of distribution b (observed during CR1912) shortly after a strong magnetic centre erupted close to the centre of the disk on 4 July during CR1911. The blue line is a polynomial fit to the quiet Sun distribution at sunspot minimum, as presented by Ortiz (2003). The vertical grey lines roughly demarcate (small) sunspots (s), micropores (m), faculae (f) and network (n), using the approximate $|B_{MDI}/\mu|$ criteria laid down by *Ortiz*, namely: pixels with radial field $|B_{MDI}/\mu| > 600$G are sunspots, $600\text{G} > |B_{MDI}/\mu| > 200$G are micropores, $200\text{G} > |B_{MDI}/\mu| > 60$G are faculae and $60\text{G} > |B_{MDI}/\mu| > 20$G are network faculae. The radial field distributions show enhanced wings when sunspots are present. For outward field ($B_{MDI}/\mu > 0$) this extends right out to 1000G for CR1912, but only to near 500G for inward field ($B_{MDI}/\mu < 0$). The distribution is also noticeably asymmetric with more flux tubes around B_{MDI}/μ of -200G than around +200G. These asymmetries make fitting the observed active region distributions problematic. Rather than using an average of the outward and inward field distributions, we here employ two different fits to the data. Model distribution 1, $n_{AR1}(B_{MDI}/\mu)/N$, shown in red, is a fit to the data for CR1912 for $0 <$

(B_{MDI}/μ) < 1000G, whereas model distribution 2, $n_{AR2}(B_{MDI}/\mu)$ /N shown in green, is a fit to the observed CR1912 distribution over the range -500G < (B_{MDI}/μ) < 0. By having significant numbers of pixels up to 1000G, model 1 has a tail which is consistent with the observed sunspots, but does not reproduce the excess pixels with (B_{MDI}/μ) around -200G which is better fitted by model 2. The two active region distributions and the quiet-Sun distribution we employ here are shown in more detail in figure 7.4.

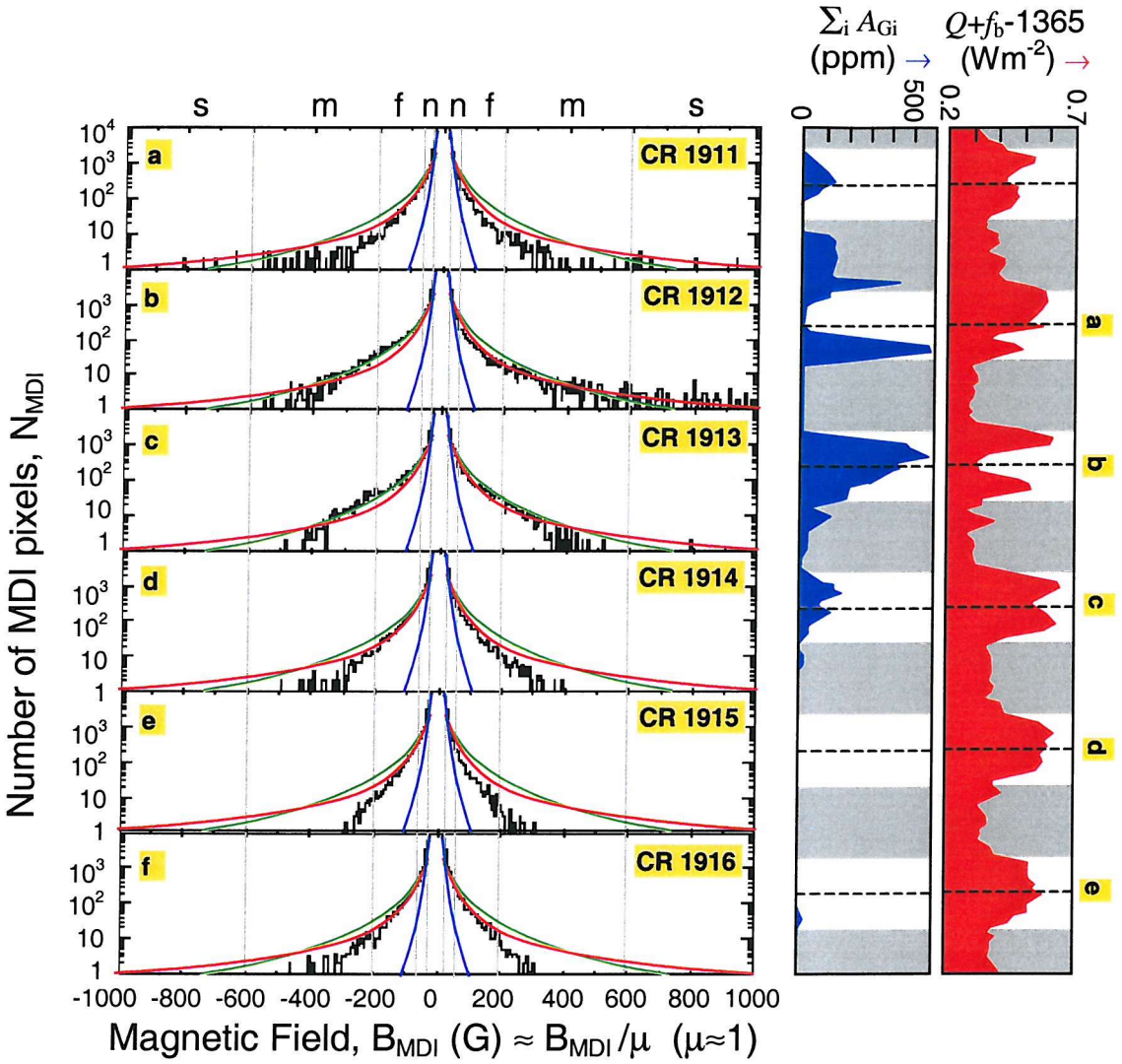


Figure 7.1. Left hand panel shows distributions of radial magnetic field observed in active region AR NOAA 7978 near the disk centre (μ≈1) by *Ortiz et al.*(2000; 2003) using full disk magnetograms from the MDI instrument of SoHO. The vertical grey lines define (small) sunspots (s), micropores (m), faculae (f) and network (n). Right hand panels show sunspot area in blue and corresponding facular brightening in red, with time running down the page.

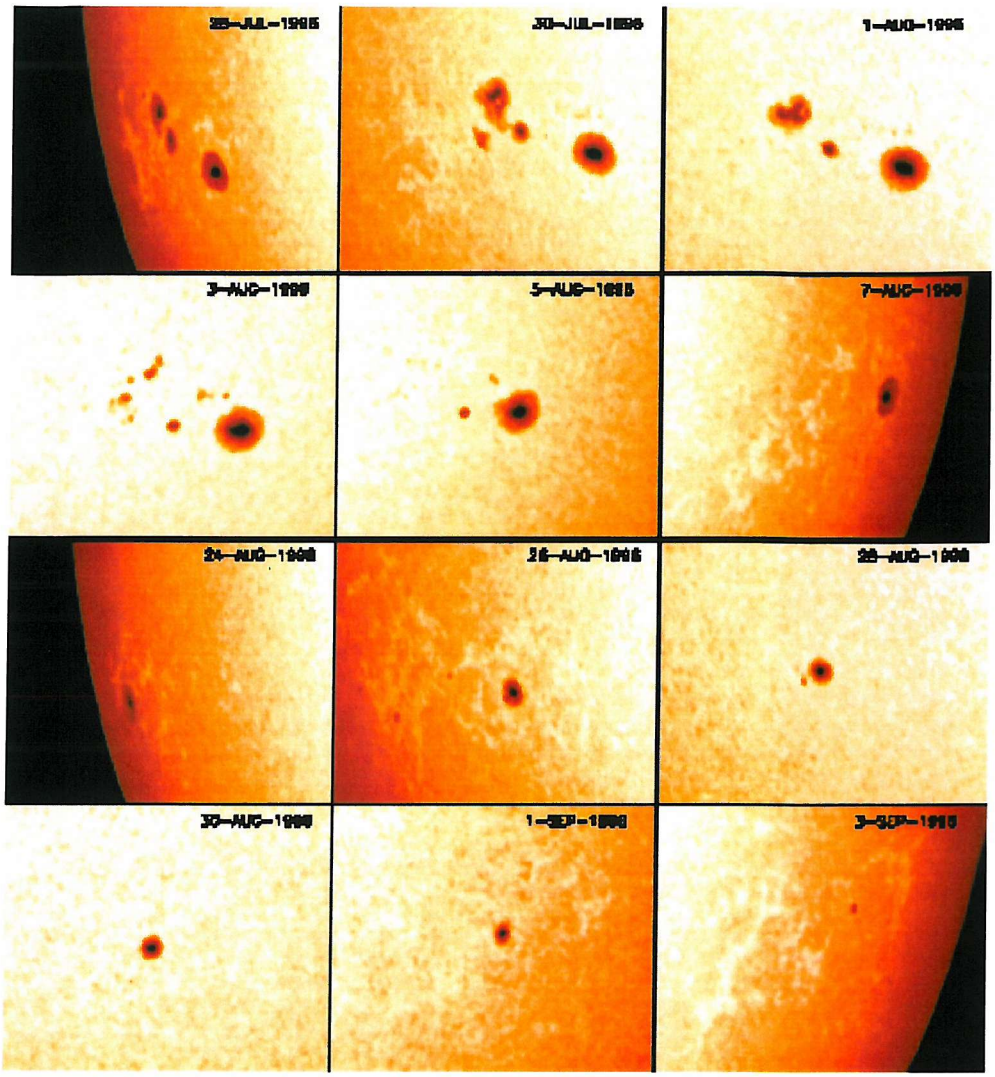


Figure 7.2. Sunspots associated with active region AR NOAA 7978, as seen in the continuum emission by SoHO/MDI during Carrington Rotations CR1912 and CR1913. Panels are images for 26 and 30 July, 1, 3, 5, 7, 24, 26, 28 and 30 August and 1 and 3 September (from *Ortiz et al.*, 2003).

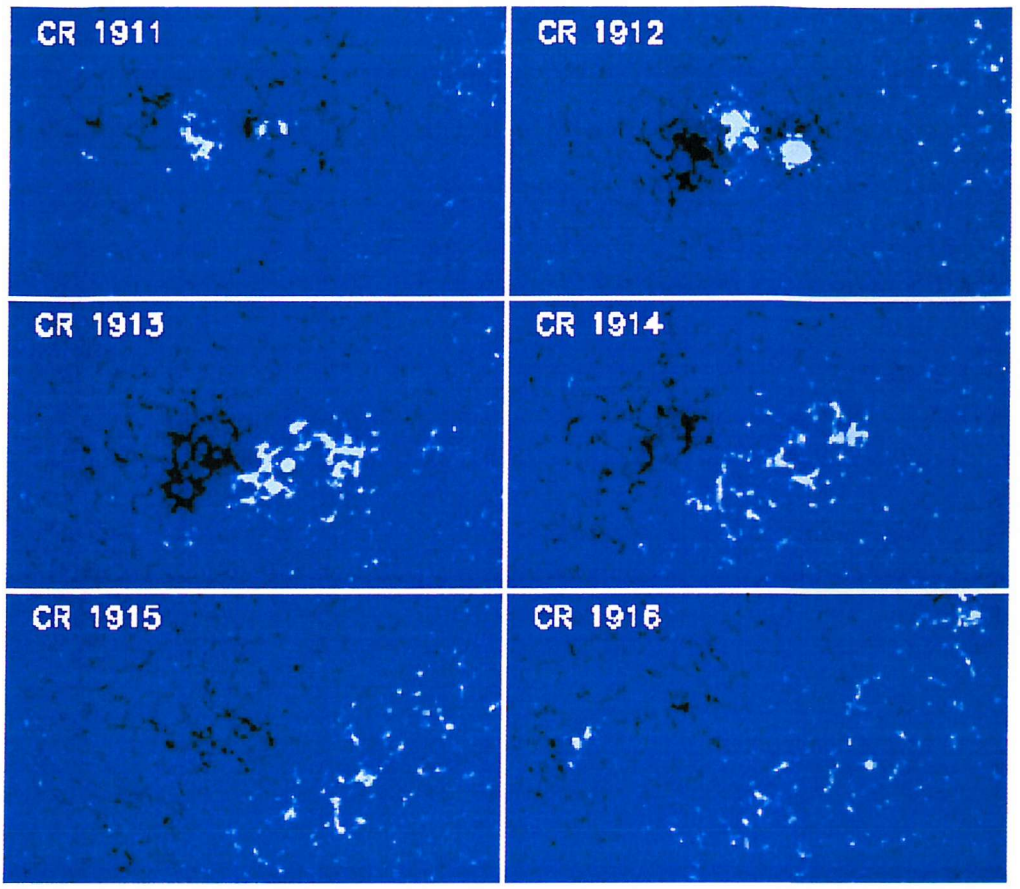


Figure 7.3. SoHO MDI magnetograms observed when active region AR NOAA 7978 was near the centre of the disk during Carrington rotations 1911-1916 (June-December 1996). Blue regions are zero field, white is outward field and black inward field. No significant sunspots were observed after the latter part of CR1913 (see figure 7.1), after which only smaller flux tubes are detected. The dispersion of the flux tubes and the increase in the area they cover can also be seen as the AR evolves (from *Ortiz et al.,* 2003).

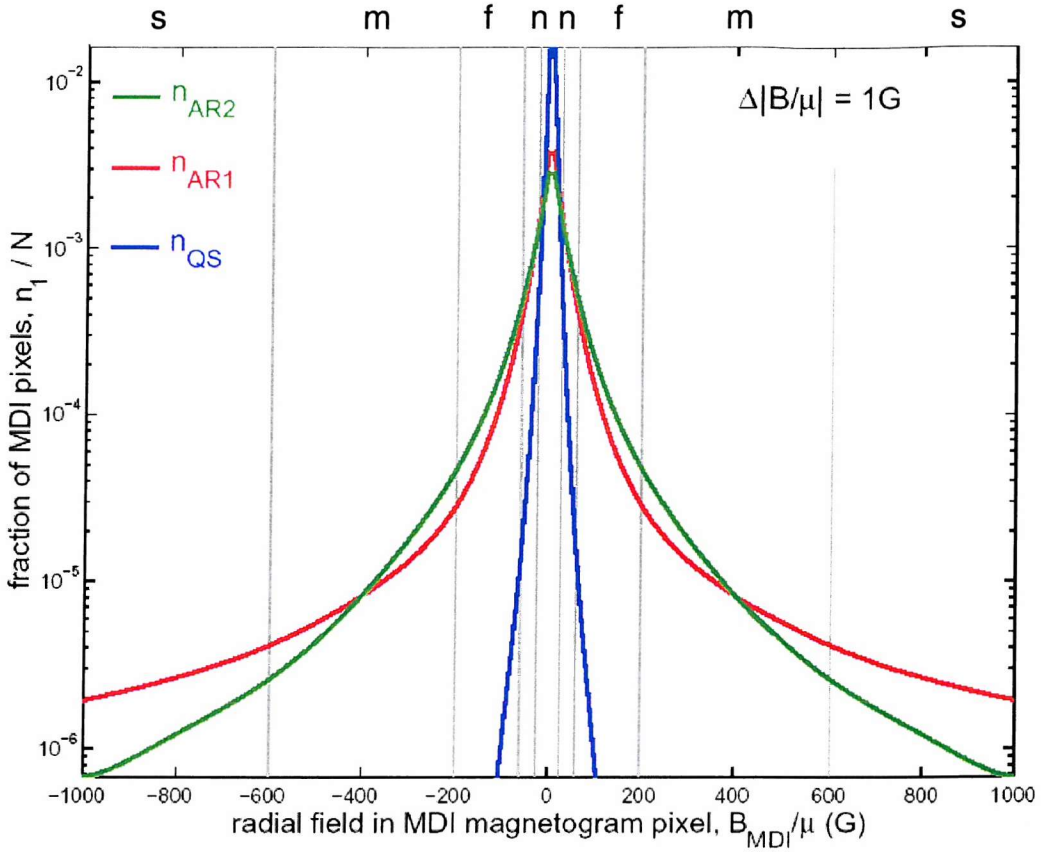


Figure 7.4. Polynomial fits to the distribution of MDI pixels with radial field B_{MDI}/μ observed for active region AR NOAA 7978 near the centre of the disc on Carrington Rotation CR1912 and (in blue) various quiet Sun regions, observed by *Ortiz et al.* [2003] between active regions and also during 1996. Because the active region was observed close to the centre of the disc ($\mu \approx 1$), in this case the radial field values are very close to the magnetogram signal values B_{MDI}. The grey lines broadly separate the different classes of magnetic flux tube, as in figure 7.1. The numbers of pixels, n_1, are for 1G bins of $|B_{MDI}/\mu|$ and summing n_1 over all such bins yields N.

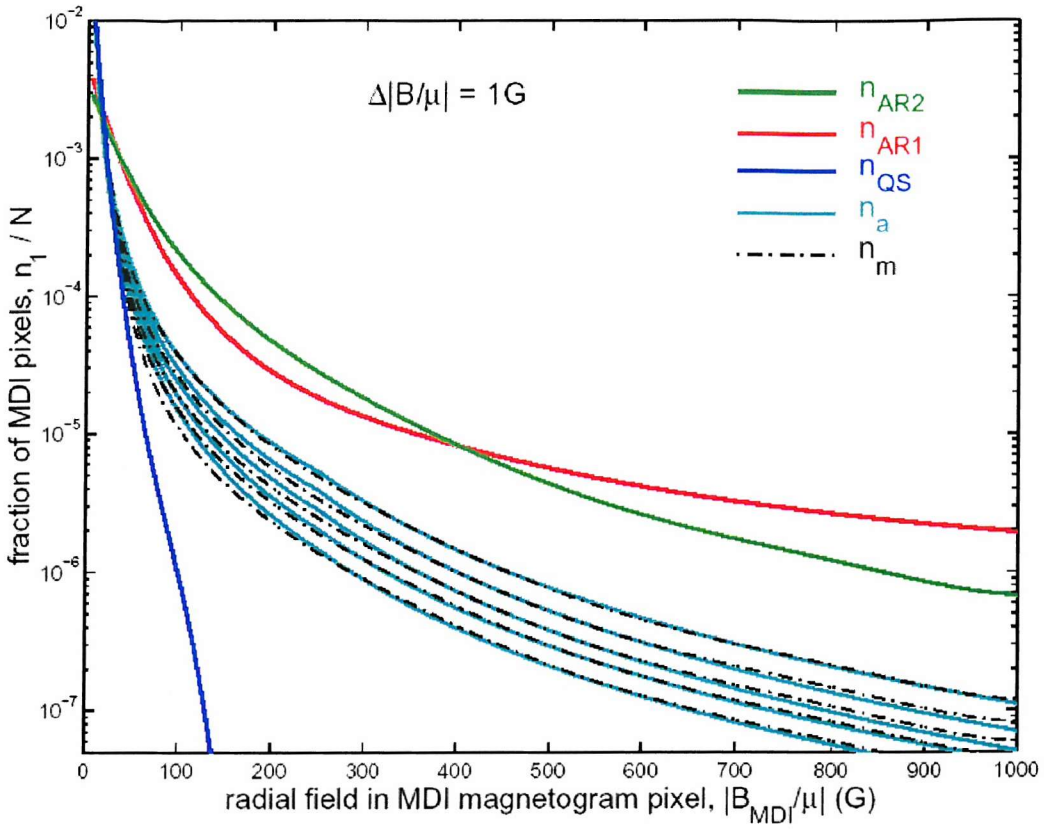


Figure 7.5. The observed annual distributions $n_a(B_{MDI}/\mu)$ from *Ortiz* et al. [2003] are here shown for 1996-2001 (cyan lines), along with the AR and QS distributions defined in figure 7.4. The dot-dashed lines are best fits $n_m(B_{MDI}/\mu)$ which are weighted combinations of n_{QS} and n_{AR2}, generated using equation 7.1.

Ortiz et al. [2003] provides an approximate definition of a faculae as having a radial field (in an MDI pixel) in the range 60-200G. With this definition, the fraction of pixels within the region in question that are faculae, F_f, equals 0.1439, 0.2162 and 0.0011 for AR model 1, AR model 2 and the quiet Sun, respectively. Thus we expect F_f to be of order 15-20% in active regions, but 0.1% in the quiet sun.

Another factor to bear in mind when evaluating these two model distributions is their ability to reproduce the observed annual distributions, $n_a(B_{MDI}/\mu)$, as given by *Ortiz et al.* [2003], who fitted observed distributions with power law variations of the form $n_a(B_{MDI}/\mu) \propto (B_{MDI}/\mu)^{-\alpha}$. Because the form of the annual distributions changes near 250G (close to where excess pixels are seen in figure 7.1 for AR NOAA 7978), *Ortiz* (2003) fitted different values of α above and below 250G. The distributions are normalised to be continuous across the 250G threshold. Because 10G bins were used to derive the distributions, we here do not extend the annual distributions below 5G and then normalise so that the sum of all pixels over the range 5-1000G is N. Table 7.1 gives the exponents α used to generate the annual distributions shown in cyan in figure 7.1.

Table 7.1. Exponents for computing annual distributions $n_a(B_{MDI}/\mu) \propto (B_{MDI}/\mu)^{-\alpha}$

Year	α for $(B_{MDI}/\mu) \leq 250G$	α for $(B_{MDI}/\mu) \geq 250G$
1996	2.5	2.8
1997	2.4	2.8
1998	2.3	2.9
1999	2.2	2.9
2000	2.1	2.8
2001	2.1	2.8

If the Sun is thought of in terms of a three component model (namely quiet sun, sunspots and faculae), to model the small flux-tube end of the distribution shown in figure 7.5 we need only the two components, the quiet sun and faculae. The overall annual distributions would then be given by the weighted sum of the AR and QS distributions where α_{AR} is the active region filling factor of the disk

$$n_m = \alpha_{AR} \times n_{AR} + (1 - \alpha_{AR}) \times n_{QS} \qquad (7.1)$$

The form of the model AR distribution 1, n_{AR1}, shown in figure 7.5 means that it cannot be combined with the quiet sun distribution using equation (7.1) to match the observed annual distributions. Specifically a third population, giving more pixels around 200G must be added. On the other hand, model AR distribution 2, n_{AR2}, also shown in figure 7.5 can be combined with the quiet sun distribution using equation (7.1) and generates the good fits shown by the dashed black lines in figure 7.5.

Figure 7.6 shows how the best fits in figure 7.5 were obtained. The root mean square deviation of the observed annual distribution n_a from the model distributions n_m, obtained from equation 7.1, is shown as a function of the AR filling factor α_{AR} for the various years. It can be seen that clear minima are found, but these are different for each year, except 2000 and 2001, which have almost identical distributions (see Table 7.1). The best-fit AR filling factor increases with time because these distributions cover the rising phase of the solar cycle. The points in figure 7.7 show the best-fit AR filling factors α_{AR} derived in figure 7.6 as a function of the corresponding annual means of sunspot group surface area, A_G, as derived from the Greenwich/USAF data and shown in figure 7.7 as a ratio of the area of a solar hemisphere, A_{SH}. The line is the best-fit linear regression fit is:

$$\alpha_{AR} = 0.0453 + 75.5957\left(\frac{A_G}{A_{SH}}\right) \qquad (7.2)$$

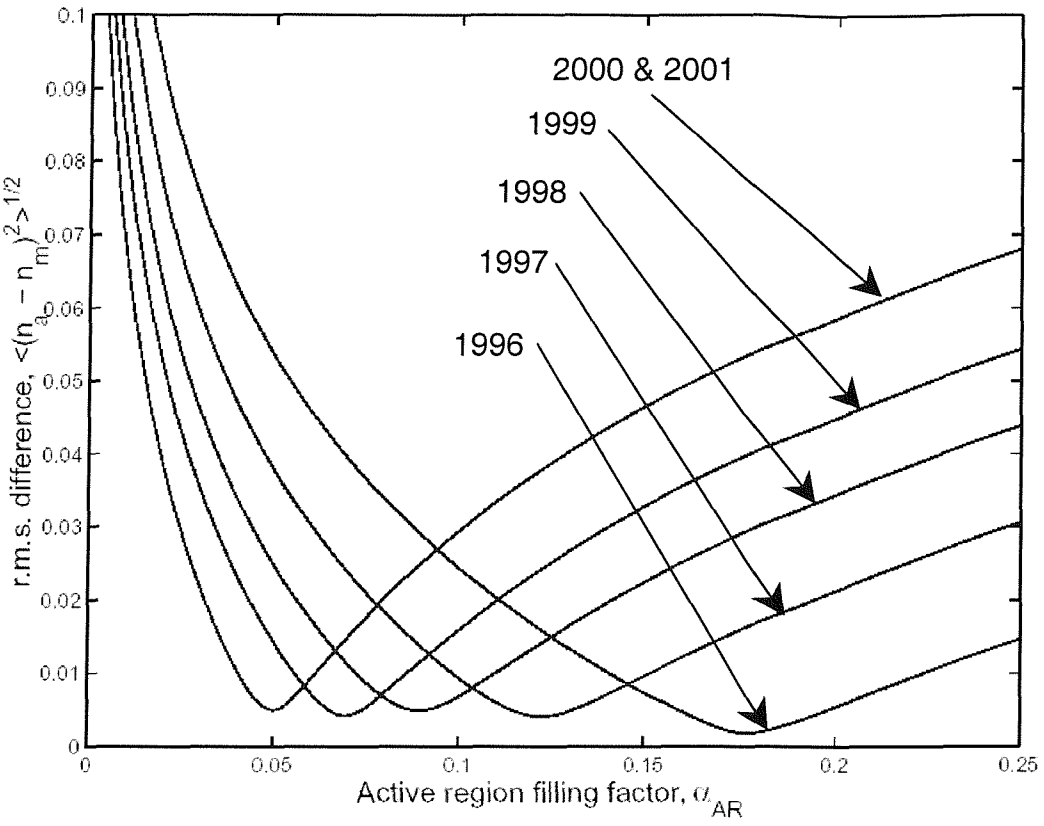


Figure 7.6. The r.m.s. deviation of observed and fitted annual distribution of radial field values, n_a and n_m respectively, as a function of the active region filling factor α_{AR}, for 1996 to 2001 (note that n_a is the same for 2000 and 2001). The minima define the best fits shown in figure 7.7.

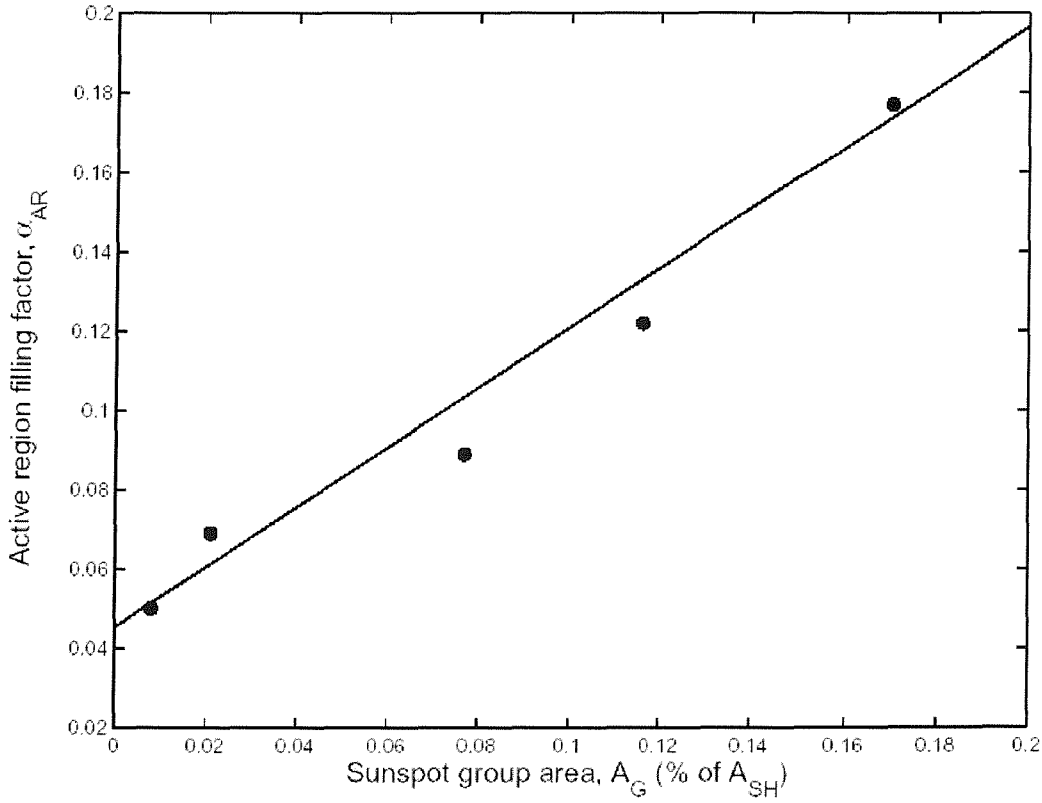


Figure 7.7. The best-fit AR filling factors α_{AR} derived in figure 7.6 as a function of the corresponding annual means of sunspot group surface area, A_G, as derived from the Greenwich/Mount Wilson data and shown as a ratio of the area of a solar hemisphere, A_{SH}. The line is the best-fit linear regression fit given by equation (7.2).

7.3. Contrasts derived from the magnetic field distributions.

Using the equations of *Ortiz et al.* [2002], we can compute the average facular contrast at a given μ for active regions and the quiet Sun, using the distributions of radial B values as shown in figure 7.4. If a given radial field (B_{MDI}/μ) gives a contrast C_{MDI} and has a filling factor f_{MDI} (= n/N where n is the number of pixels showing radial field (B_{MDI}/μ) in an area A_D of the visible disk and N is the total number of pixels in the same area A_D). The effect on irradiance is proportional to the product $f_{MDI}C_{MDI}$ and this is shown, as a function of B_{MDI}/μ and μ, in figures 7.8, 7.9 and 7.10 for, respectively, the model active region (AR) distributions, n_{AR1} and n_{AR2}, and the quiet Sun (QS) distribution n_{QS}.

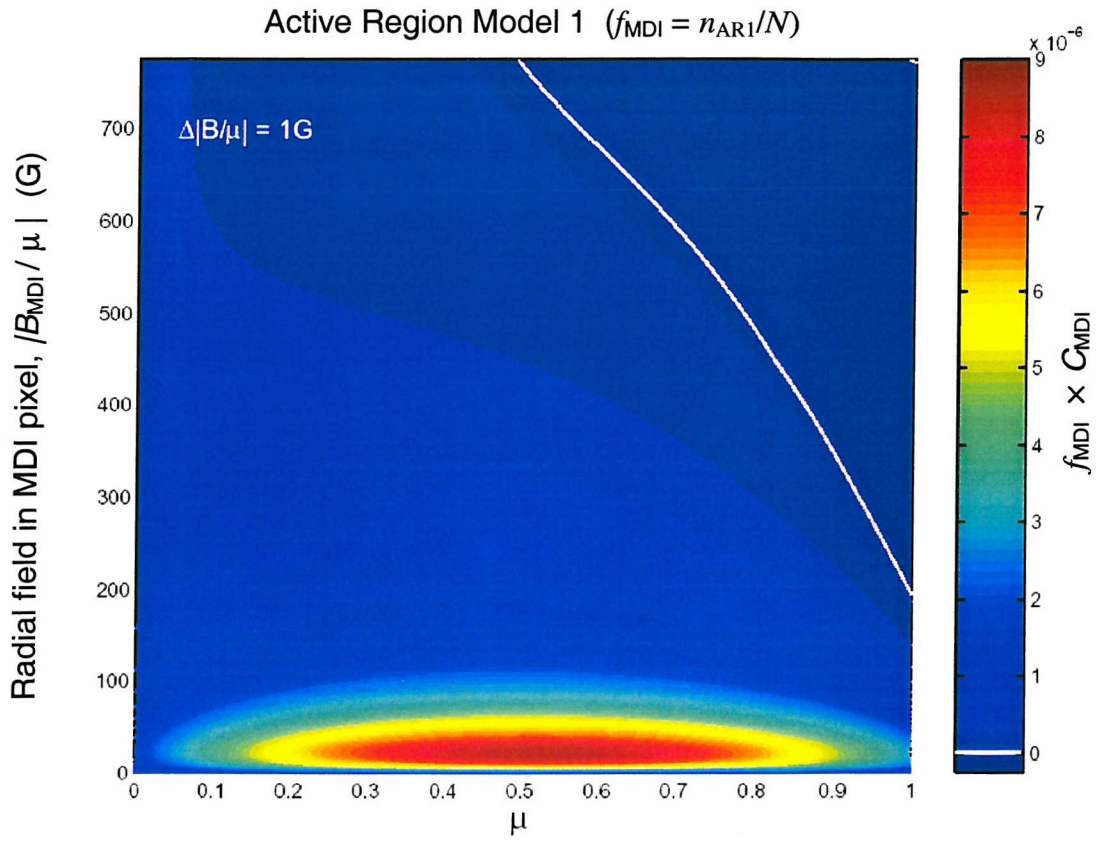


Figure 7.8. The contrasts of active region pixels, normalised by the occurrence of MDI pixels of the radial field strength (B_{MDI}/μ) in question, $f_{MDI}C_{MDI}$ (f_{MDI} = n/N where n is the number of pixels in a given area A_D of the visible disk with B_{MDI}/μ in 1G bins and N is the total number of such pixels in the area A_D) – plotted here as a function of B_{MDI}/μ and μ. This is for the model AR distribution 1, n = n_{AR1}, shown in red in figures 7.1, 7.4 and 7.5.

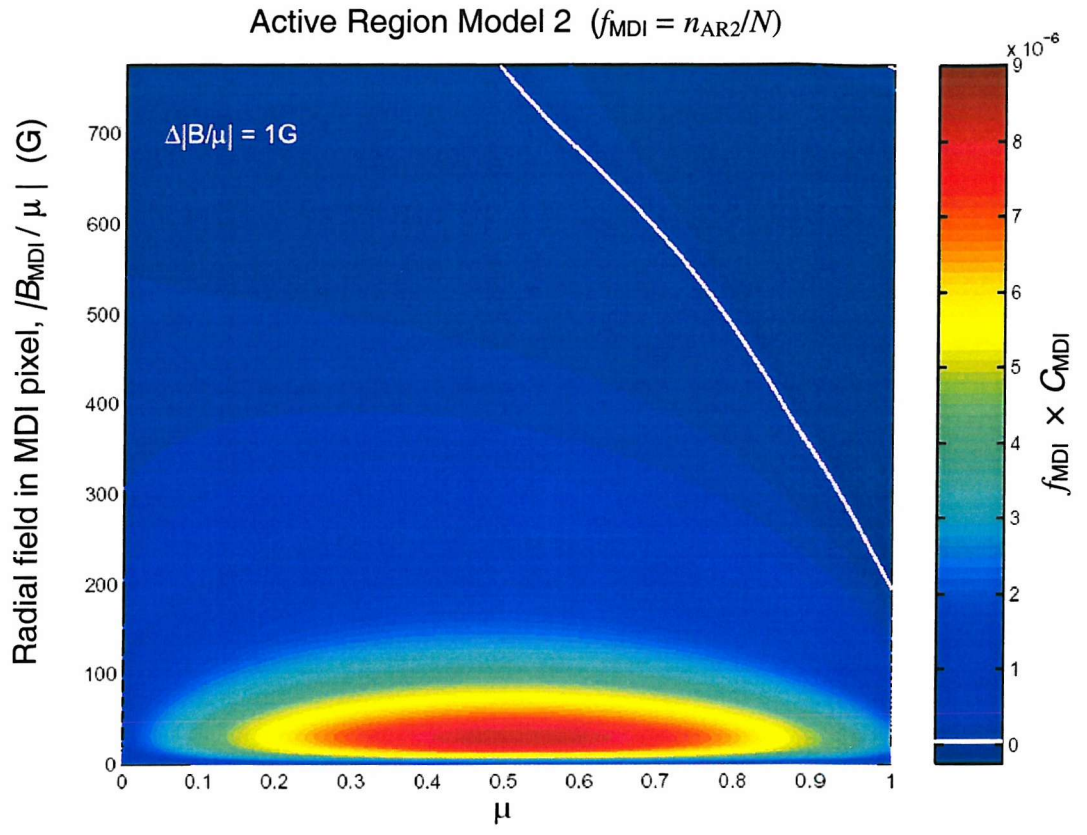


Figure 7.9. Same as figure 7.8, for the model AR distribution 2, n_{AR2}, shown in green in figures 7.1, 7.4 and 7.5.

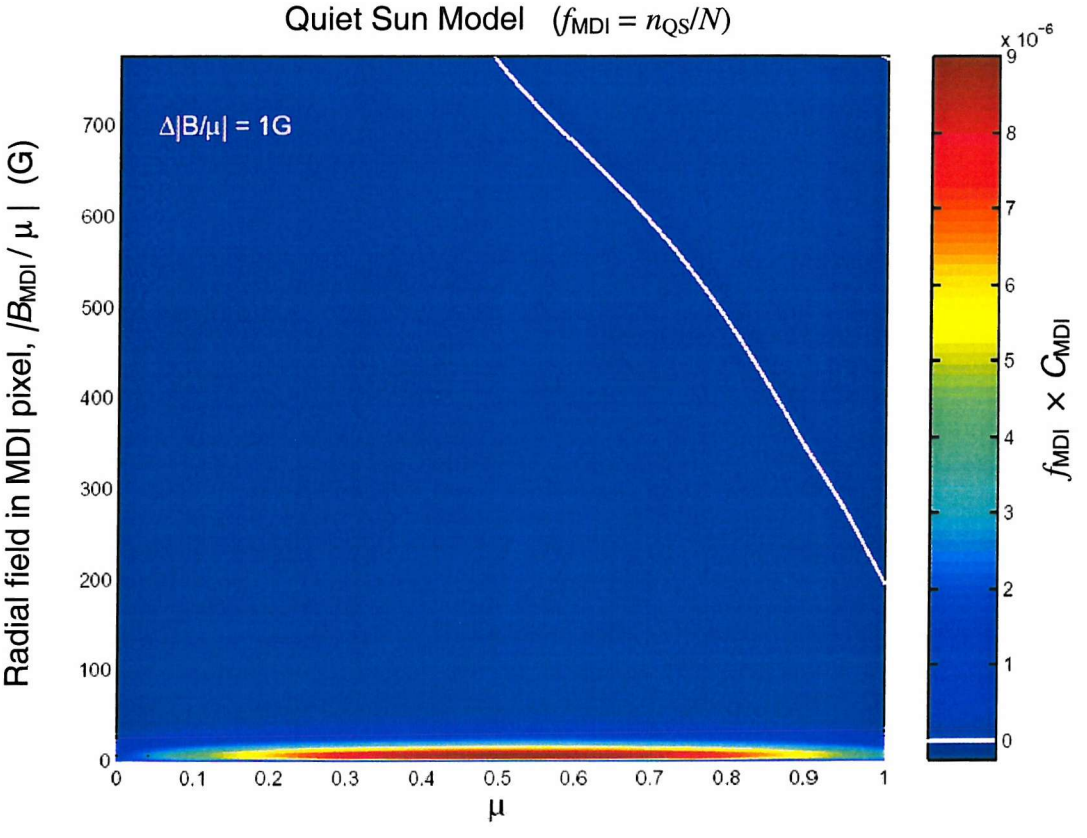


Figure 7.10. Same as figure 7.8, for the solar-minimum quiet Sun distribution, n_{QS}, shown in blue in figures 1, 7.4 and 7.5.

It can be seen that the brightening effect of active region flux tubes is mainly in the range 200G > $|B_{MDI}/\mu|$ > 60G, which we here characterise as faculae in active regions or ephemeral regions. Brightening by flux tubes in the quiet sun is restricted to small flux tubes (less than about 60G), but is also significant. The latter will be dominated by network faculae and ephemeral flux tubes. As can be seen the region of greatest brightening is to be found in the region of 0.4 to 0.6 μ. This is somewhat away from the solar limb, where features achieve the greatest increase in contrast relative to the limb darkening.

In figure 7.11 we plot the averages of the distribution of contrasts at a given μ, $\langle C_{MDI} \rangle$, as a function of μ (i.e. the values in figures 7.8 – 7.10 have been averaged over the $|B_{MDI}/\mu|$ dimension). These are shown by the red, green and blue lines for AR model 1, AR model 2 and QS, respectively. This figure clearly shows the dependence of the contrast of bright active region features on the position of the feature on the solar disk. If we take a faculae that rotates into the field of view at the equator (μ=0) and then moves across the solar surface, we will find that its contrast relative to the quiet sun, will increase to a maximum at μ = 0.45, before decreasing sharply towards the centre of the solar disk. As the faculae then moves towards the eastern solar limb this effect will occur in reverse.

As discussed in chapter 5, as an active region moves across the solar surface it's μ value will change being smallest at the solar limb and increasing to a maximum when it is closest to the centre of the solar disc. This variation in μ according to latitude is shown in figure 5.5, and it can be seen that active regions at high solar latitudes will only ever achieve relatively small μ values, since they remain close to the solar limb at all times of their transit, whereas active regions which occur at low solar latitudes will span the full range of μ values. This is important, since as was found in Chapter 3, the latitudes at which sunspots have occurred has increased steadily over the past 100 years, and so average μ values are expected to have fallen. This can be seen in figure 7.13, which shows monthly averages of the sunspot group μ values since 1874, and it can seen that μ values, as expected, have fallen significantly since 1900. If facular μ values have changed in a similar fashion then it could have an impact upon the contrast of these active region faculae, an effect investigated later in

this chapter. Note that figure 7.13 uses the sunspot group data composite developed in chapter 5.

The intensity of an area of the Sun is proportional to the product of the mean contrast $<C>$ and the limb darkening function L_D (see equation 7.4 in section 7.4). This product is shown as a function of the disc position parameter μ, for the same three model distributions in figure 7.12. The shape of the red and green curves is very similar, such that the ratio $<(<C_{AR2}>L_D)>_D$ / $<(<C_{AR1}>L_D)>_D$ is approximately constant over all but the smallest and largest μ (near the disc limb and centre) the mean value is 1.256 and the standard deviation 0.05. Figure 7.12 shows that when the effect of limb darkening is taken into account, the shapes of the intensity curves, as a function of μ, is somewhat different to those of the contrast. The position at which maximum intensity is at $\mu = 0.58$ (nearer the disc centre than the peak contrasts which were around $\mu = 0.45$).

Figure 7.12 is also interesting as it shows that bright features such as faculae are not brightest at the limb, but at a distance somewhat closer to the centre of the solar disk. This is important since studies use the core to limb ratio to evaluate facular brightening, whereas it may be better to use a centre to $\mu = 0.5$ relation to optimise these studies.

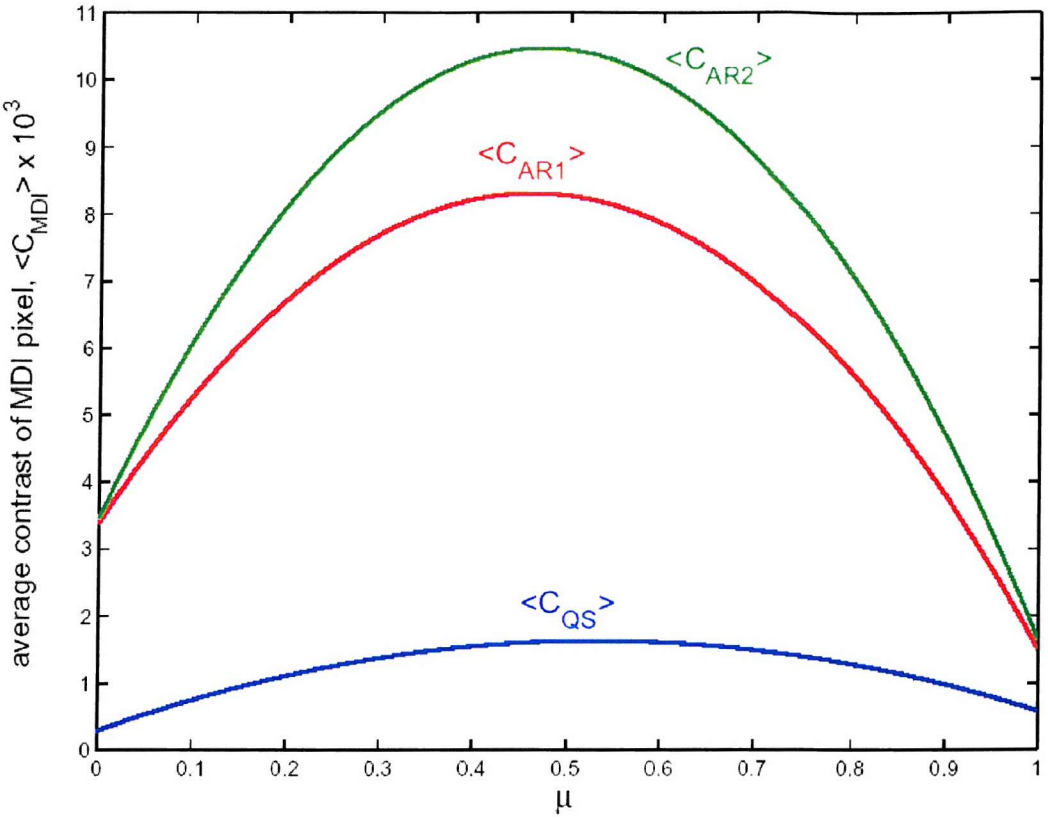


Figure 7.11. The average contrast of MDI-sized pixels at a given disc position parameter μ, for radial field distributions: active region model 1 (red, $<C_{AR1}>$) active region model 2 (green, $<C_{AR2}>$), solar minimum quiet Sun distribution (blue, $<C_{QS}>$).

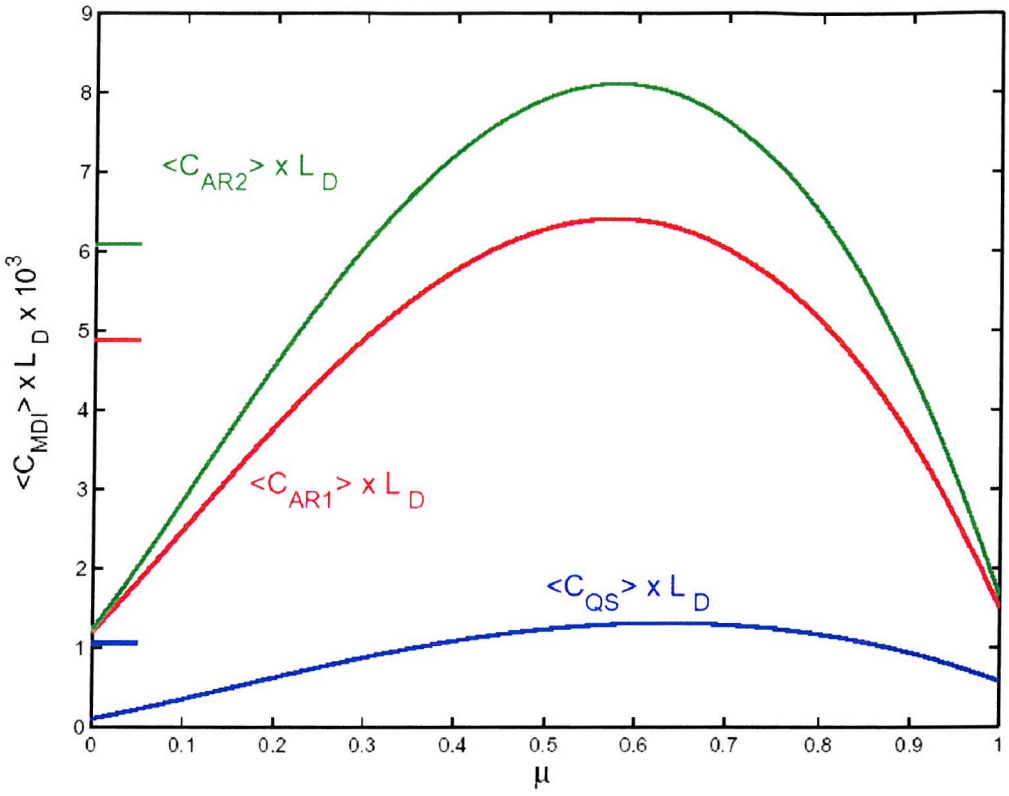


Figure 7.12. Average contrasts multiplied by the limb-darkening factor (the product being proportional to intensity) at a given disc position parameter μ, for radial field distributions: active region model 1 (red, $\langle C_{AR1}\rangle L_D$) active region model 2 (green, $\langle C_{AR2}\rangle L_D$), solar minimum quiet Sun distribution (blue, $\langle C_{QS}\rangle L_D$). The short horizontal lines on the left give the corresponding disc-averaged values $\langle(\langle C_{AR1}\rangle L_D)\rangle_D = 4.874 \times 10^{-3}$; $\langle(\langle C_{AR2}\rangle L_D)\rangle_D = 6.086\times 10^{-3}$; and $\langle(\langle C_{QS}\rangle L_D)\rangle_D = 1.057\times 10^{-3}$.

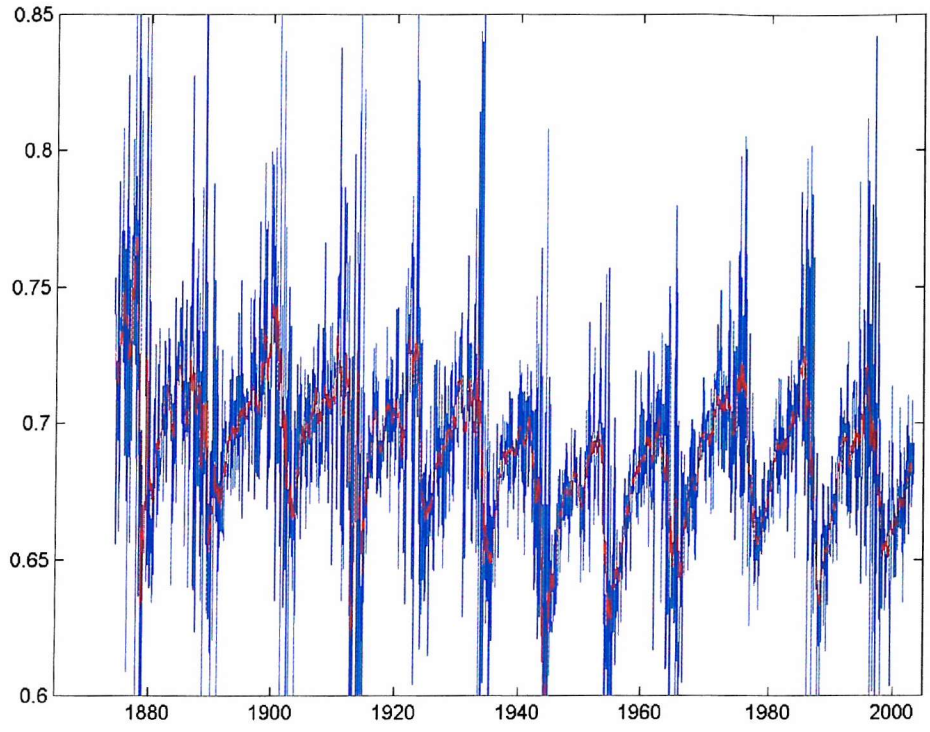


Figure 7.13. Long-term variation of sunspot μ values for the period 1874 – 2003 from the composite sunspot group data series developed in Chapter 5. Blue: daily values, red: monthly averages.

7.4. Calculating facular Brightening.

To reconstruct the brightening effect of small flux tubes in active regions, consider the effect of an active region of surface area A_{AR}. This will occupy an area $A'_{AR} = \mu A_{AR}$ on the visible disc of the Sun. The region has an average contrast $<C_{AR}>$, relative to the intensity of a magnetic-field free pixel at the same μ, the latter showing the limb-darkening factor L_D which, like $<C_{AR}>$, is a function of μ. Were the same region to be a "quiet Sun" photosphere, its contrast would be $<C_{QS}>$, relative to the intensity of a magnetic-field free pixels at the same μ.

The "filling factor" of this region on the solar disc is:

$$\alpha_{AR} = \frac{A'_{AR}}{\pi r_s^2} = \frac{N_{AR}}{N_S} \tag{7.3}$$

where N_{AR} is the number of MDI pixels within the active region and N_S is the total number of MDI pixels in the solar disc. From the geometry we can compute the surface area of an MDI pixel as a function of μ.

The change in total solar irradiance due to the facular brightening of one pixel is:

$$\delta f_{b1} = I_{AR} - I_{QS} = I_o\left(\langle C_{AR}\rangle + 1\right) - I_o\left(\langle C_{QS}\rangle + 1\right) = [I_o]_{(\mu=1)} L_D\left(\langle C_{AR}\rangle - \langle C_{QS}\rangle\right) \tag{7.4}$$

where I_o is the field-free intensity of that pixel and $[I_o]_{(\mu=1)}$ is the field-free intensity of a pixel at the centre of the solar disc (μ=1).

Summing over all pixels in the group, which is assumed small enough that $<C_{AR}>$ and L_D are approximately constant:

$$\delta f_{bi} = [I_o]_{(\mu=1)} \sum_{i=1}^{N_{AR}} \left[\langle C_{AR}\rangle - \langle C_{QS}\rangle\right]_i L_{Di} = N_{AR}[I_o]_{(\mu=1)} \left[\langle C_{AR}\rangle - \langle C_{QS}\rangle\right]_i L_{Di} \tag{7.5}$$

If we sum over all N_S pixels for a magnetic-free Sun, we derive a total field-free solar irradiance:

$$Q_o = [I_o]_{(\mu=1)} \sum_{i=1}^{N_s} L_{Di} = N_s [I_o]_{(\mu=1)} \langle L_D \rangle_D \tag{7.6}$$

where $<L_D>_D$ is the disc-averaged limb-darkening factor, which in section 4.8.2 we showed was equal to 0.8478 for the wavelength at which MDI operates.

Substituting (7.6) and (7.3) into (7.5) yields

$$\delta f_{bi} = \alpha_{ARi} \left\{ \frac{Q_o}{\langle L_D \rangle_D} \right\} \left[\langle C_{AR} \rangle - \langle C_{QS} \rangle \right]_i L_{Di} \tag{7.7}$$

If we then sum over all N active regions present on the solar disc at any one time we get the total contribution of active regions to facular brightening:

$$f_{ba} = \sum_{i=1}^{N} \delta f_{bi} = \left[\frac{Q_o}{\langle L_D \rangle_D} \right] \sum_{i=1}^{N} \alpha_{ARi} \left[\langle C_{AR} \rangle - \langle C_{QS} \rangle \right]_i L_{Di} \tag{7.8}$$

We can also apply an equivalent equation to the present day quiet Sun and look at the brightness increase of the quiet sun at solar minimum, relative to a magnetically free Sun, using the average contrasts $<C_{QS}>$. If we assume that the quiet Sun is featureless (so $<C_{QS}>$ is, like L_D, a function of μ only) we replace the discrete sum over all active regions with an integral over the whole disc (equation 54, chapter4). In this case, the brightening predicted is due to network faculae and any other small flux tubes that are present during recent solar minima, f_{bno}.

$$f_{bno} = \left\{ \frac{Q_o}{\langle L_D \rangle_D} \right\} \times 2 \int_0^1 L_D \langle C_{QS} \rangle \mu d\mu = \left\{ \frac{Q_o}{\langle L_D \rangle_D} \right\} \times \left\langle \left(\langle C_{QS} \rangle L_D \right) \right\rangle_D \tag{7.9}$$

Figure 7.12 shows that $<(<C_{QS}>L_D)>_D = 1.057\times10^{-3}$ and figure 4.9 shows that $<L_D>_D = 0.8478$. Hence:

$$\frac{f_{bno}}{Q_o} = \frac{\left(1.057\times10^{-3}\right)}{0.8478} = 1.247\times10^{-3} \qquad (7.10)$$

Allowing for sunspot darkening (quantified by the PSI, P_{SI}), the surface field-free total solar irradiance Q_o, facular brightening (f_{bn} in the network and f_{ba} in active regions, where f_{bn} has a present-day solar minimum value f_{bno} and a solar cycle variation Δf_{bn}), the general expression for total solar irradiance is:

$$I_{TS} = Q_o + f_{bno} + \Delta f_{bn} + f_{ba} - P_{SI} \qquad (7.11)$$

From equation (7.8) we can compute f_{ba}/Q_o, and we later will show that $(\Delta f_{bn} + f_{ba}) = s f_{ba}$, where $s = 1.20 \pm 0.06$, is the slope of the best-fit linear regression (figure 7.28), which gives the best fit to observations. From (7.11) and (7.10) this yields

$$Q_o = \frac{\{I_{TS} + P_{SI}\}}{\left\{1 + 1.247\times10^{-3} + s\dfrac{(f_{ba})}{Q_o}\right\}} \qquad (7.12)$$

The green line in figure 7.14 shows the results of applying equation (7.12) to annual values, the blue line gives $(Q_o + f_{bno})$ from equation (7.10). The black line is the observed $(Q_o + f_b) = (Q_o + f_{ba} + f_{bno} + \Delta f_{bn}) = (I_{TS} + P_{SI})$. The red line is $(Q_o + f_{bno} + \Delta f_{bn})$ obtained from (11) by subtracting the f_{ba} variation derived below from $(I_{TS} + P_{SI})$. It can be seen that the different years give a very consistent value of Q_o the mean value of which is 1363.8 ± 0.1 Wm^{-2}. From (10), this yields $f_{bno} = 1.701$ Wm^{-2} for the interval covered by figure 7.14..

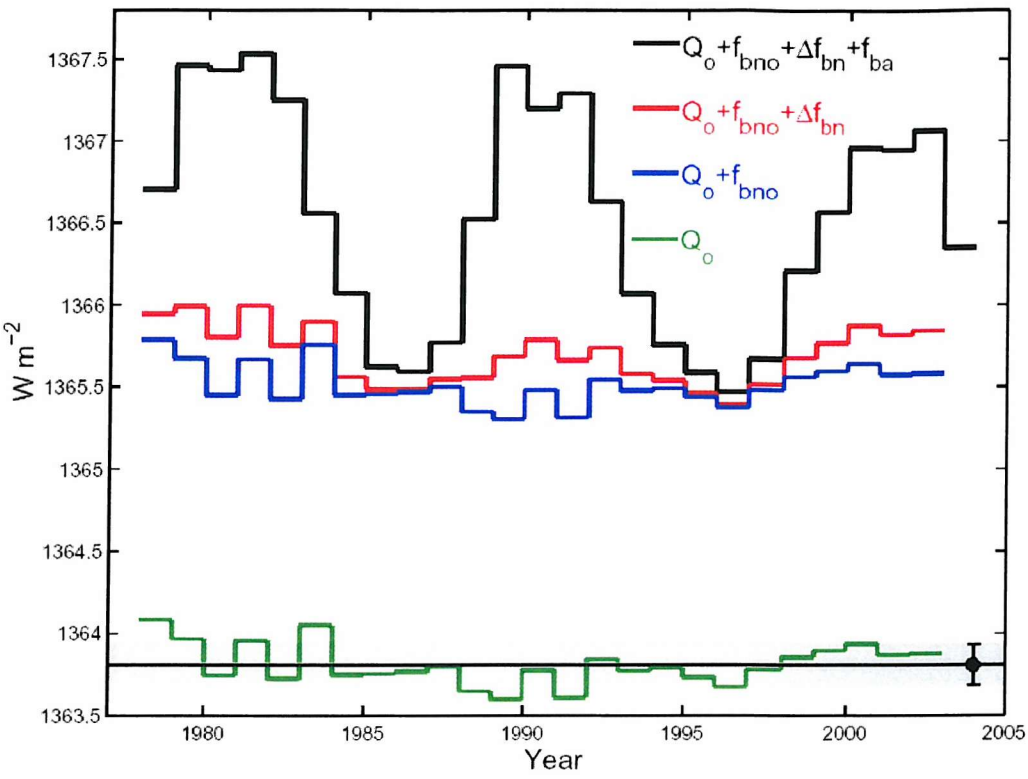


Figure 7.14. Histograms of annual means of: (black) the observed $(Q_o + f_b) = (I_{TS} + P_{SI})$; (red) $(Q_o + f_{bno} + \Delta f_{bn})$: (green) Q_o and (blue) $(Q_o + f_{bno})$.

This estimate of the irradiance of the fully quiet, magnetic-free Sun of Q_o = 1363.8 ± 0.1 Wm^{-2} is highly significant. It means that, the network facular brightening at recent sunspot minimum was small (f_{bno} = 1.701 Wm^{-2}). Thus considering the removal of all solar-minimum surface magnetic field and the associated brightening, sets an upper limit to the possible irradiance that can result (from surface emissitivity changes at least) and sets a lower limit to the fully-quiet value, Q_o, of 1363.7 Wm^{-2} for the Maunder minimum. In fact, $I_{TS} = Q_o$ in the Maunder minimum is unlikely to be valid as the ^{10}Be cosmogenic isotope record through the Maunder minimum shows a continuing solar cycle variation [Beer et al., 1990], indicating flux emergence continued.

Of the various irradiance reconstructions discussed in the introduction (chapter 1), all violate this value of Q_o (by giving $I_{TS} < Q_o$ in the Maunder minimum), including those by *Solanki and Fligge* [1999], *Hoyt and Schatten* [1993], *Lean et al.* [1995] and *Lean* [2000] (and hence that by *Lockwood and Stamper*, [1999] and the reconstruction presented in Chapter 6). This can be seen in figure 7.15, which shows

the irradiance reconstructions mentioned above, with respect to the fully quiet sun irradiance value. The closest to this limit are the reconstructions by *Lean* [2000] and *Lockwood and Stamper* [1999], which correspond to I_{TS} in the Maunder minimum of 1363.5 Wm^{-2} and although this is close to the field free irradiance derived here, it is still lower than Q_o, even allowing for the uncertainty. Even our reconstruction based on the open flux–TSI correlation *Solanki et al.* [1999] model (chapter 6), gives TSI that is too low by 0.7 Wm^{-2}. One way that these values can be resolved with the considerations given here is if there are subsurface effects (alpha or beta effects) in the convection zone, which add to the irradiance change. However, this conclusion is based on the contrasts developed by *Ortiz et al.* [2002] from MDI pixel resolution. If on smaller scales there is cancelling flux as discussed by *Kivrova et al.* [2002] then there could be additional brightening of the quiet sun over the magnetic-free Sun and our Q_o estimate could be too large (and f_{bno} is too small). The scatter in the *Ortiz et al.* contrasts is of order ± 25% at low $|B_{MDI}/\mu|$ (see figure 4.10). Applying this error to the average contrast and substituting into equation (7.12) yields an uncertainty in Q_o of ± 0.4 Wm^{-2}. The upper limit of 1364.2 Wm^{-2} is still smaller than the base level of 1365.3 Wm^{-2} which *Fröhlich* [2003a] arrives at when he subtracts all facular brightening associated with best-fit Mg II index proxies. The lower limit of 1363.4 Wm^{-2} is consistent with the *Lean* [2000a] and the *Lockwood and Stamper* [1999] reconstructions, as well as that presented in chapter 6. Thus these reconstructions could be explained by the 'hidden flux' in the MDI magnetograms. However the *Lean et al.* [1995], *Solanki and Fligge* [1999] and *Hoyt and Schatten* [1993] reconstructions fall below even this lower limit and so appear to have too large an amplitude of long-term drift even allowing for the potential hidden flux effect.

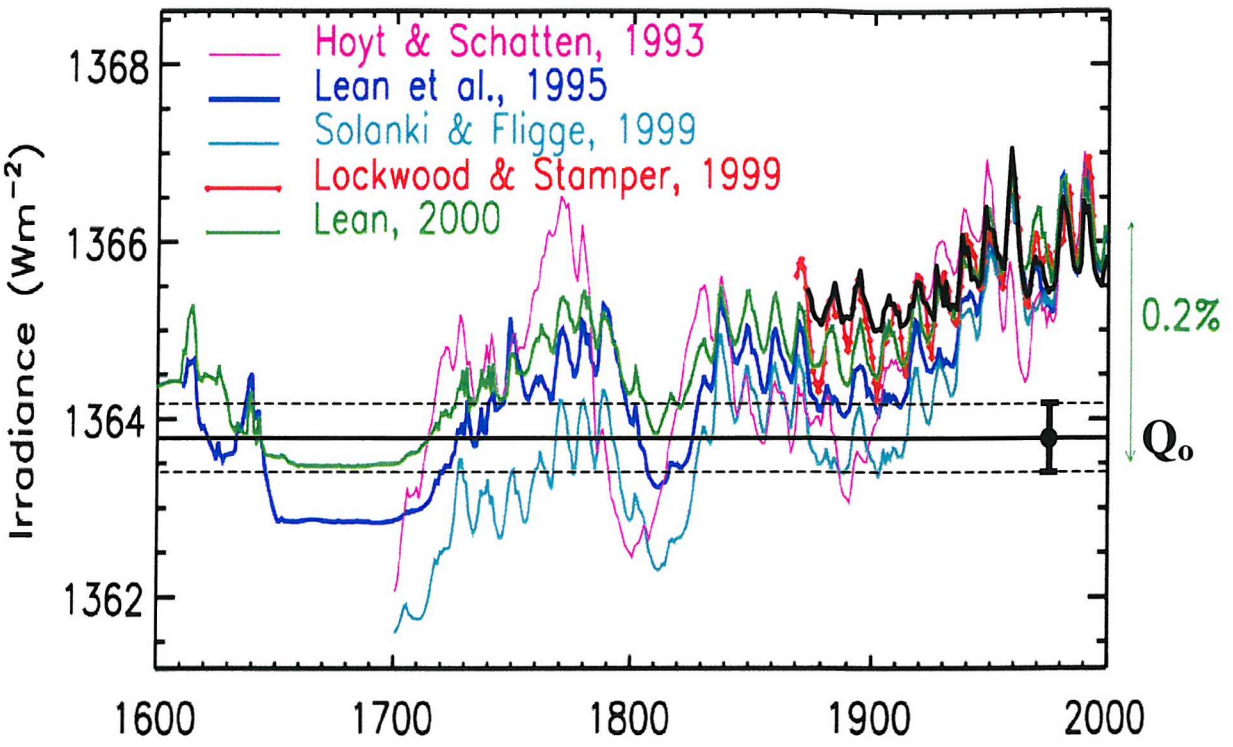


Figure 7.15. Plot showing various TSI reconstructions, and the Q_o cut off point that I have established through this work.

7.5 Application to Sunspot Group Data and Filling Factors

From the combined Greenwich/USSR/Mt. Wilson dataset, we know the surface area of all sunspot groups A_G and their μ values (see section 5) from 1874 to the present day. In order to use the sunspot group data as a proxy for active region facular area (and thereby model distributions of solar magnetic fields in this interval), we need to know how the sunspot group area relates to the area A_{AR} (and hence the filling factor α_{AR}) to which our active region model radial field distributions apply.

We here use two methods to determine the ratio A_G/A_{AR}. In section 7.5.1 we employ the known relationship sunspot group area and the facular area A_f [*Chapman et al.*, 2001] and this can be applied to either of the two models radial field distributions for active regions n_{AR1} and n_{AR2}. The second method (section 7.5.2) uses the fits n_m, to the observed annual distributions of radial field n_a, as discussed in

section 7.2; however, this can only be used with the distribution n_{AR2} because n_{AR1} cannot be combined with n_{QS} to generate a n_m that are good fits to n_a.

7.5.1. Method 1: using facular area, A_f.

The regression fit between monthly means of A_f and the sunspot group area from the Greenwich/Mt. Wilson data, A_G is shown in figure 7.16. As discussed in section 2.2.3, the facular data we have used comes from the Cartesian Full-Disk Telescope No.1 (CFDT1) at the San Fernando observatory, (http://www.csun.edu/sfo/instruments.html). The threshold used to define a faculum in the San Fernando data is a K-line contrast exceeding 4.8%/μ (*Chapman et al.*, 2001, and references therein).

The 1/μ term is an approximate way to reduce the detection bias caused by limb brightening of the facular pixels on K line contrast images. Note that the *Ortiz* contrasts are relative to the limb profile and so include an equivalent effect, but the centre-to-limb variations involved are not the same. This is one reason why it is not straightforward to find out precisely what *Ortiz et al* contrast threshold would correspond to the *Chapman* faculum threshold. Other reasons are that the spatial resolution is different and that the wavelengths are different (indeed, Ortiz et al. use the continuum around 676.8 nm whereas *Chapman* uses the 393.4 nm CaII K line emission). These differences present a problem in equating the using the Ortiz et al magnetic field distributions, and the resulting distributions of contrasts, in association with the facular areas observed the San Fernando observatory. The problem is that we do not know what fraction F_f of the active region pixels, used to compute mean contrast, are sufficiently bright that they would be classed as faculae in the San Fernando data.

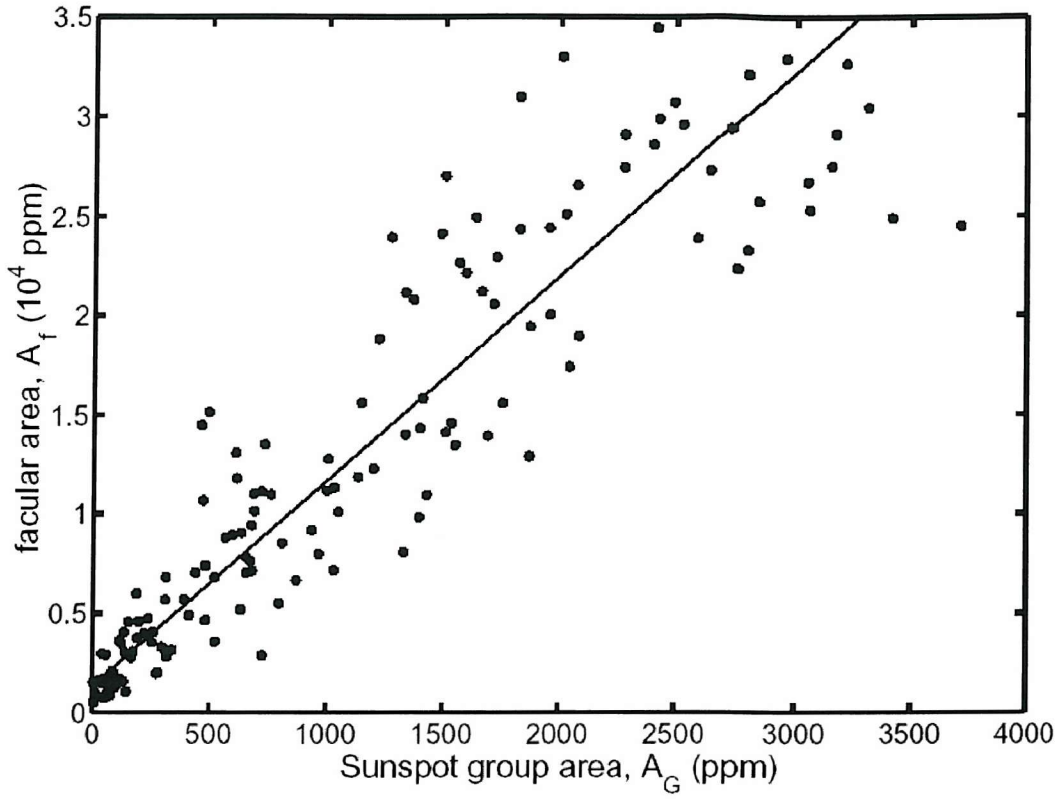


Figure 7.16. Linear regression of the facular surface area A_f (in ppm of a solar hemisphere) against sunspot group area from the Greenwich/USAF catalogue, A_G (also in ppm per solar hemisphere, horizontal axis). The linear regression line [A_f in ppm] = s_{fG} [A_G in ppm] + i_{fG} has slope s_{fG} = 10.30 and intercept i_{fG}. = 0.1357. The correlation coefficient is 0.917, which is significant at the 99.97% level.

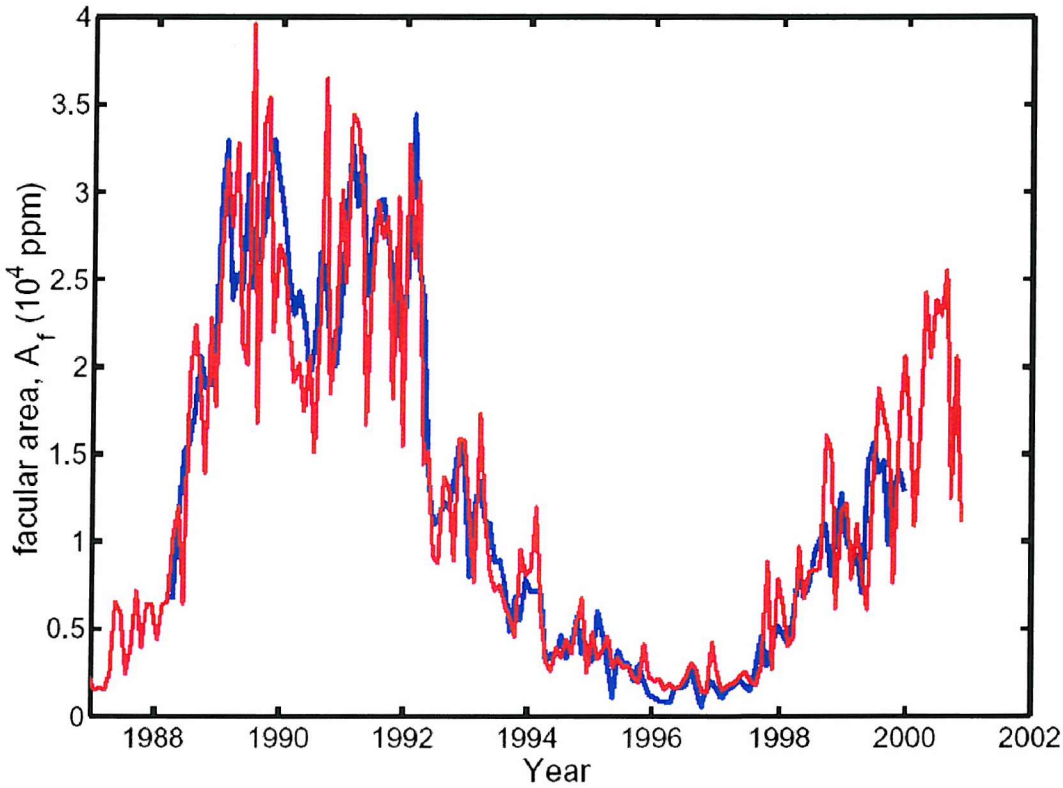


Figure 7.17. Temporal variations of the of the facular surface area A_f (in ppm of a solar hemisphere) against sunspot group area from the Greenwich/Mount Wilson catalogue, A_G (also in ppm per solar hemisphere, horizontal axis): A_G has been fitted to the A_f scale using the regression shown in figure 7.16.

The area A_f on the surface classed as facular corresponds to an underlying active region area of $A_{AR} = A_f / F_f$. The area of the corresponding sunspot group A_G is related by the regression fit shown in figure 7.16.

$$\frac{A_{AR}}{A_G} = \frac{s_{fG}}{F_f} \tag{7.19}$$

We here omit the intercept i_{fG} because this represents the background faculae present when A_G goes to zero, rather than the active region brightening. Thus to find the ratio A_{AR}/A_G we need to evaluate F_f. *Ortiz et al.* [2003] suggests that the range of radial fields 60G < (B_{MDI}/μ) <200G broadly corresponds to faculae. As mentioned above, if we adopt this condition, we find that the fraction of active-region MDI pixels, which meet this criterion is 0.144 and 0.216 for AR models 1 and 2, respectively. However, this can be considered no more than a rough guide. Here we initially use $F_f = 1$ and then use the results to evaluate the required fraction.

For the μ value of the ith active region present, μ_i and the surface area of the ith active region A_{ARi}, we can evaluate its area on the visible disc $A'_i = A_{ARi}\mu_i$ and apply the average active-region contrast $C_A(\mu_i)$ to give a facular brightening contribution. The number of MDI pixels in the area A'_i is:

$$N_i = \frac{A'_i}{A_{MDI}(\mu_i)} = \frac{A_{ARi}\mu_i}{A_{MDI}(\mu_i)} = \frac{\{A_{Gi}s_{fG}\mu_i\}}{\{F_f A_{MDI}(\mu_i)\}} \tag{7.20}$$

where $A_{MDI}(\mu_i)$ is the area of the disc covered by one MDI pixel at μ_i. The filling factor of this region on the solar disc is given by equation (7.3), $\alpha_{ARi} = N_i / N_S$, where N_S is the total number of MDI pixels on the solar disc. From (7.7):

$$\delta(f_{ba})_i = \left\{\frac{N_i}{N_s}\right\}\left\{\frac{Q_o}{\langle L_{DG}\rangle_D}\right\}\langle C_{AR}(\mu_i)\rangle L_D(\mu_i) \tag{7.21}$$

and we then sum over all N sunspot groups present to get the total active region facular brightening:

$$f_{ba} = \sum_{i=1}^{N} \delta(f_{ba})_i = \left\{ \frac{Q_o}{\langle L_{DG} \rangle_D} \right\} \sum_{i=1}^{N} \left\{ \frac{N_{Gi}}{N_S} \right\} \langle C_{AR}(\mu_i) \rangle L_D(\mu_i) \qquad (7.22)$$

equations (20) and (22) allow us to construct (for a given value of F_f) the active region facular brightening f_{ba} from the sunspot group data composite for all times back to 1874.

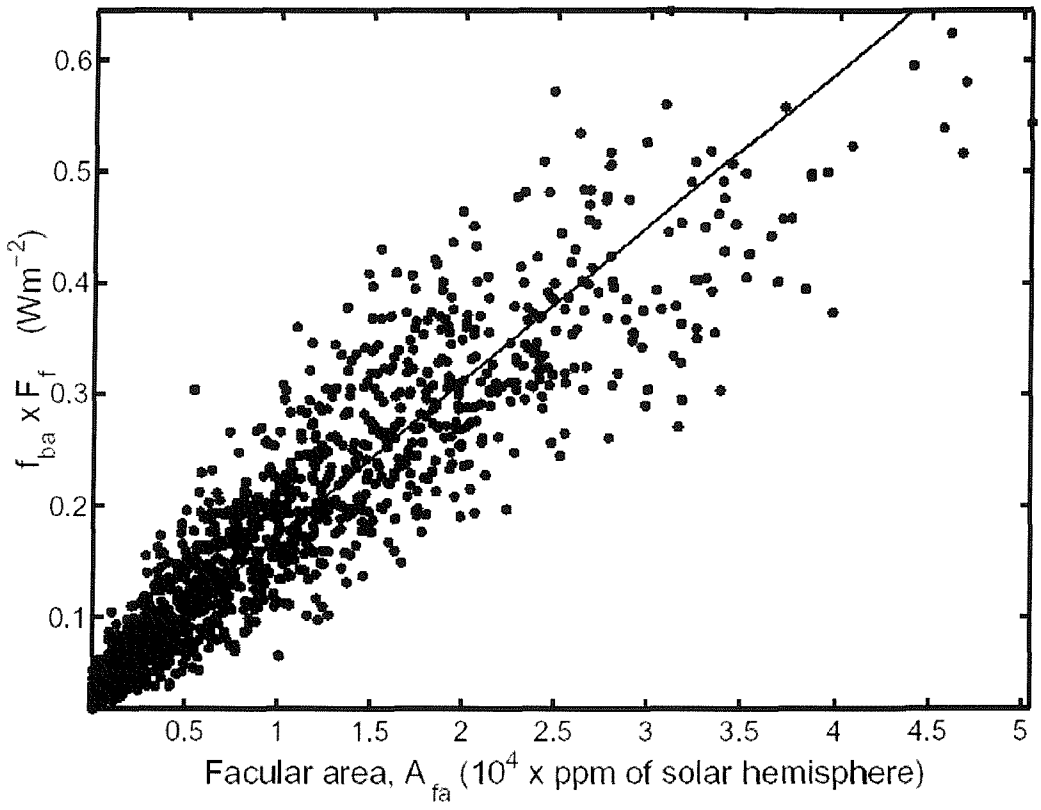


Figure 7.18. Scatter plot of monthly means of the total active-region facular brightening modelled from the sunspot group data composite $f_{ba} = \Sigma_i \delta(f_{ba})_i$ (where i refers to the i^{th} sunspot group) for unit area factor F_f, as a function of the input facular area $A_{fa} = \Sigma_i A_{fi} = s_{fG} \Sigma_i A_{Gi}$.

Initially we here use a facular area fraction F_f of unity. This means that we use an area of active regions that is F_f times the real values. Therefore the active region brightening predicted is the real value, f_{ba}, multiplied by F_f. The results, monthly means for 1874 to 2001, are plotted in figure 7.18 as a function of the active-region

facular area employed $A_{fa} = \sum_i A_{fi} = s_{fG} \sum_i A_{Gi}$. The contrasts used to derive figure 7.18 are for model 1 of the active-region radial field distribution, n_{AR1}. The correlation coefficient is very high indeed, peaking at 0.9681 with the facular brightening lagged by 1 ± 1 month. This lag comes from the tendency for faculae to remain after the sunspot group has dispersed (see figure 7.1). Figure 7.18 shows that the active region area has the dominant role in determining the brightening effect. However there is scatter, which largely arises from the different positions of the active regions on the solar disc. The slope of the best regression fit in figure 7.18 is:

$$\frac{\Delta\left[f_{ba}F_f\right]}{\Delta A_{fa}} = 0.1390 \text{ Wm}^{-2} \text{ (ppm of a SH)}^{-1} \qquad (7.23)$$

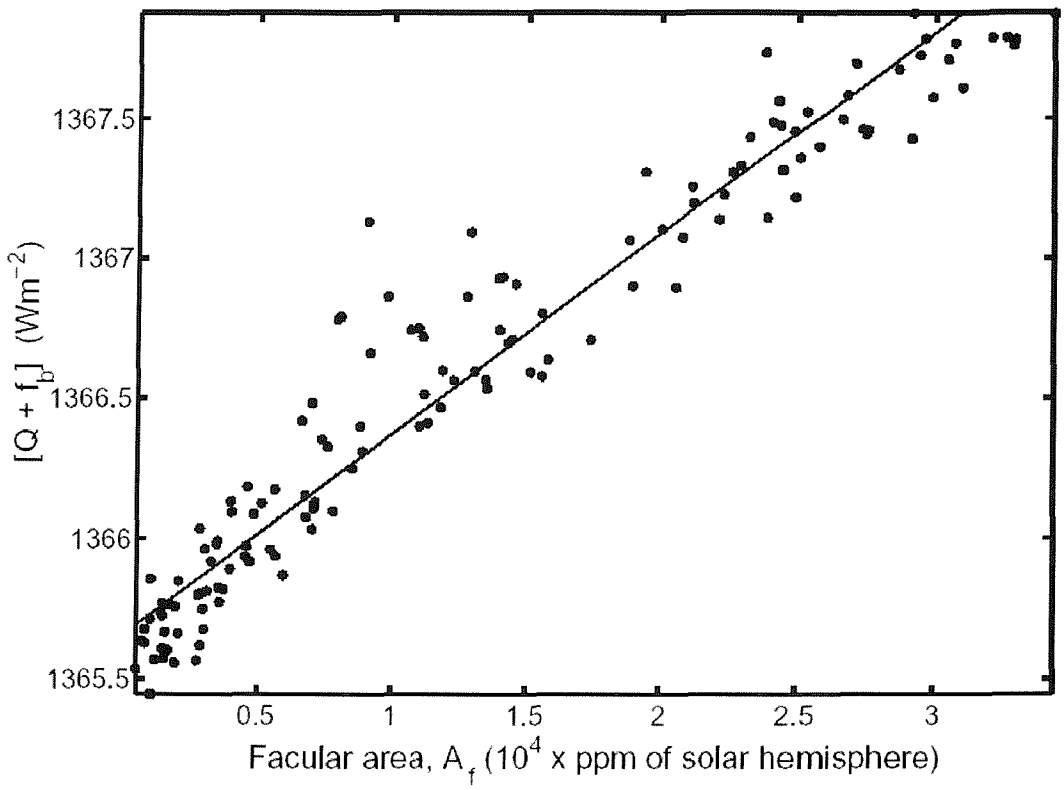


Figure 7.19. Scatter plot of monthly means of the total surface facular area A_f (as a fraction of a solar hemisphere) from the San Fernando observatory as a function of $(Q_o + f_b) = (I_{TS} + P_{SI})$, where I_{TS} is the TSI from the PMOD composite variation and P_{SI} is the PSI). The correlation coefficient is $r = 0.9657$, which is significant at the 75% level and explains $r^2 = 93.3$ % of the variation. The solid line is the best linear regression fit $(Q + f_b) = 0.7174$ [A_f in 10^4 ppm]+1365.65.

Figure 7.18 can be compared with the equivalent plot from observations over the last solar cycle, as shown in figure 7.19. The total facular brightening is f_b and, from equation (7.10), equals ($I_{TS} + P_{SI} - Q_{min}$), where I_{TS} is the total solar irradiance (TSI), P_{SI} is the photometric sunspot index (PSI) and Q_{min} is the modern-day quiet sun irradiance (= $Q_o + f_{bno}$). We here compare this estimate of the total facular brightening with the total surface area A_f covered by faculae, as observed by CFDT1/San Fernando. The observed values show a nearly linear relationship, as demonstrated in figure 7.19. The modern-day quiet-Sun irradiance Q_{min}, can be obtained from the observed A_f =0 intercept, and the linear regression yields a value of 1365.65 Wm^{-2}. The slope is:

$$\frac{\Delta f_b}{\Delta A_f} = 0.7174 = \frac{\Delta f_{ba}}{\Delta A_{fa}} \text{ Wm}^{-2} \text{ (ppm of a SH)}^{-1} \qquad (7.24)$$

The RHS of equation (7.24) assumes that the brightness-to-area ratio for network faculae is the same as for active region faculae. Even if there is a difference between the ratios for these two classes of faculae, equation (7.24) will still be approximately valid because both Δf_b and ΔA_f are dominated by the effect of active regions [*Walton et al.*, 2003]. From (7.23) and (7.24), $F_f = 0.1938$. This is within the 15-20% range derived from the approximate faculum definition of *Ortiz et al.* [2003].

7.5.2. Method 2: using the annual distributions of radial field, n_a.

Figure 7.5 shows that the observed annual distributions $n_a(B_{MDI}/\mu)$ derived by *Ortiz et al.* [2003] for 1996-2001 can be modelled well by, $n_m(B_{MDI}/\mu)$, which are weighted combinations of n_{QS} and n_{AR2}, generated using equation (7.1). To calculate the facular brightening equation (7.2) can be used directly to give the active region filling factor $\alpha_{ARi} = \{N_{Gi} / N_S\}$ corresponding to a sunspot group area A_{Gi} and thus can be employed in equation (7.21) to give the facular brightening caused by the group, $\delta(f_{ba})_i$. However, it is instructive to evaluate the corresponding factor F_f. Again neglecting faculae present when $A_G = 0$, and combining equations (7.2) and (7.19)

$$\alpha_{AR} = 75.596\left(\frac{A_G}{A_{SH}}\right) = 75.596\frac{\left(A_{AR}F_f\right)}{\left(A_{SH}s_G\right)} = \frac{A'}{\left(\pi r_s^2\right)} = \frac{A_{AR}}{A_{SH}} \qquad (7.25)$$

yielding $F_f = 0.1363$. Thus the best fits to the annual distributions $n_a(B_{MDI}/\mu)$, requires an F_f value which is comparable with, but slightly smaller than that derived by method 1 from model distribution 1 (F_f = 0.1938) but is very close to that from *Ortiz*'s approximate definition of a faculum (F_f = 0.1439).

This general consistency in the derived area factors F_f gives us confidence that the area factor and brightening compared are not strongly dependent on the model distribution (and the method it is used with). Because method 2 does not use the observations of TSI and PSI in any way, being based only on the SoHO MDI field distributions, it is desirable as it is fully independent of the irradiance data. Hereafter we use this method exclusively. However, method 1 has provided a valuable check of our data and all results presented are found to be almost identical to their equivalent derived using method 1.

Using the area of faculae in active regions calculated by this method, f_{ba} can be modelled using the same equations as used by method 1.

7. 6. Sunspot and faculae positions.

In the previous section we compared facular area with the sunspot group area. If we are to apply the sunspot group data to active region faculae we must assume that the distribution of μ values for the faculae is the same as that for the sunspot groups. This assumption is verified in this section.

White light faculae measurements of position and area were made at the Greenwich observatory from 1886 – 1917. Although a distinction should be made between white light faculae and active region faculae, these records do give an indication of the distribution of active region faculae with respect to sunspots. Figures 7.20 a and 7.20 b show contour butterfly plots of sunspot and faculae latitudes, with

the contours being based upon the relative number of sunspots or white light faculae in one degree latitude bins. It can be seen that the positions of white light faculae closely match the sunspot group positions, with the butterfly distributions lying almost in identical positions. This similarity in the positions gives confidence that the sunspot μ values can be assigned to the faculae for which the areas are calculated. Thus as well as facular area we know the positions of the active region faculae, and thus facular brightening can be calculated from this calculated area and assigned μ values.

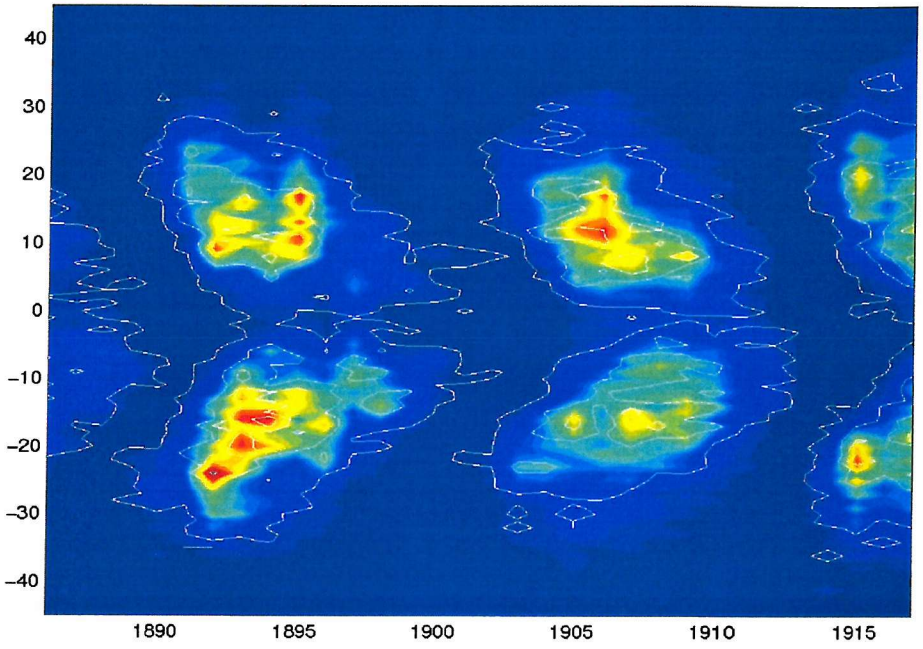


Figure 7.20(a). Contour butterfly plots of sunspot and facular heliographic latitude. Contour lines are defined by number of spots/ faculae at each latitude. For this plot the white lines denote the sunspot contours, with the coloured contours defining facular latitude.

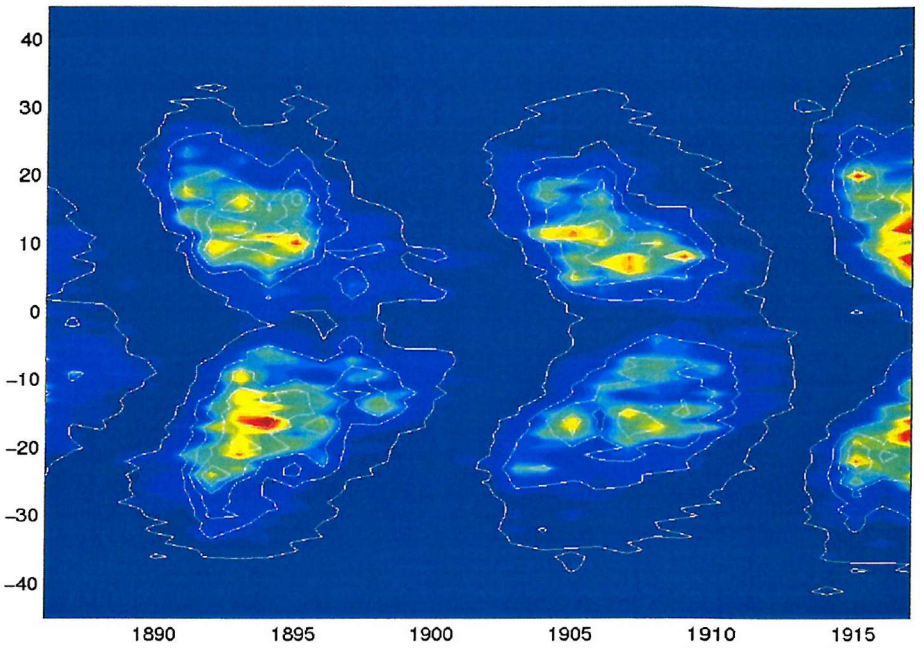


Figure 7.20(b). The same data as in part (a) but here the white lines denote the faculae contours, with the coloured contours defining sunspot latitude.

7. 7. Components of Irradiance Change.

Because using method 2 gives us an estimate of active region brightening that is fully independent of the observed I_{TS} and P_{SI}, in this section we employ the f_{ba} values derived using filling factors obtained by method 2, so that we can compare with fully independent data on $(Q_o + f_b) = (I_{TS} + P_{SI})$. Figure 7.24 compares the observed $(Q_o + f_b)$ variation with the best-fit active region facular brightening derived from the sunspot area. Figure 7.1 shows that the facular brightening associated with an active region shows a more gradual rise and fall than the sunspot group area and we here use a 3-point running mean to allow for this effect. Figure 7.1 also shows that the facular brightening effect lags behind the sunspot group area, and we need to allow for such a lag also.

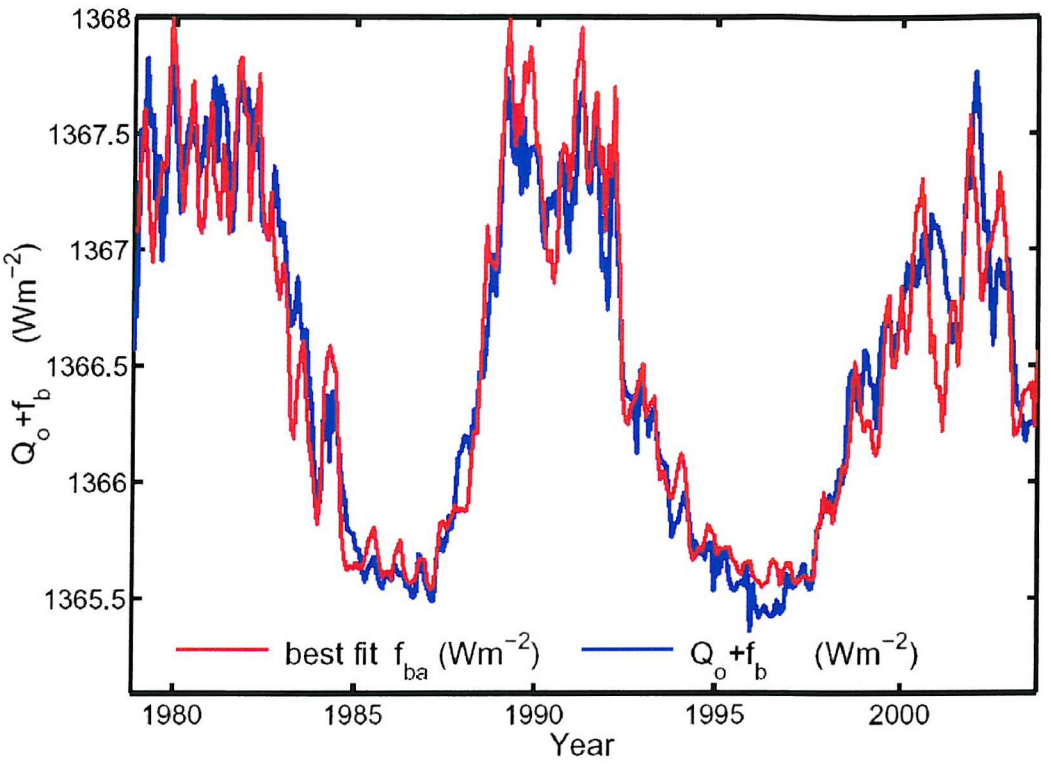


Figure 7.21. Linear regression fit of the reconstructed active-region facular brightening f_{ba} (from sunspot group data, red), fitted to the observed brightening, $Q_o + f_b = I_{TS} + P_{SI}$ (blue). Three-month running means of f_{ba} have been used with the optimum lag of one month (see figure 7.22). The f_{ba} data have been scaled using the best-fit linear regression line shown in figure 7.23.

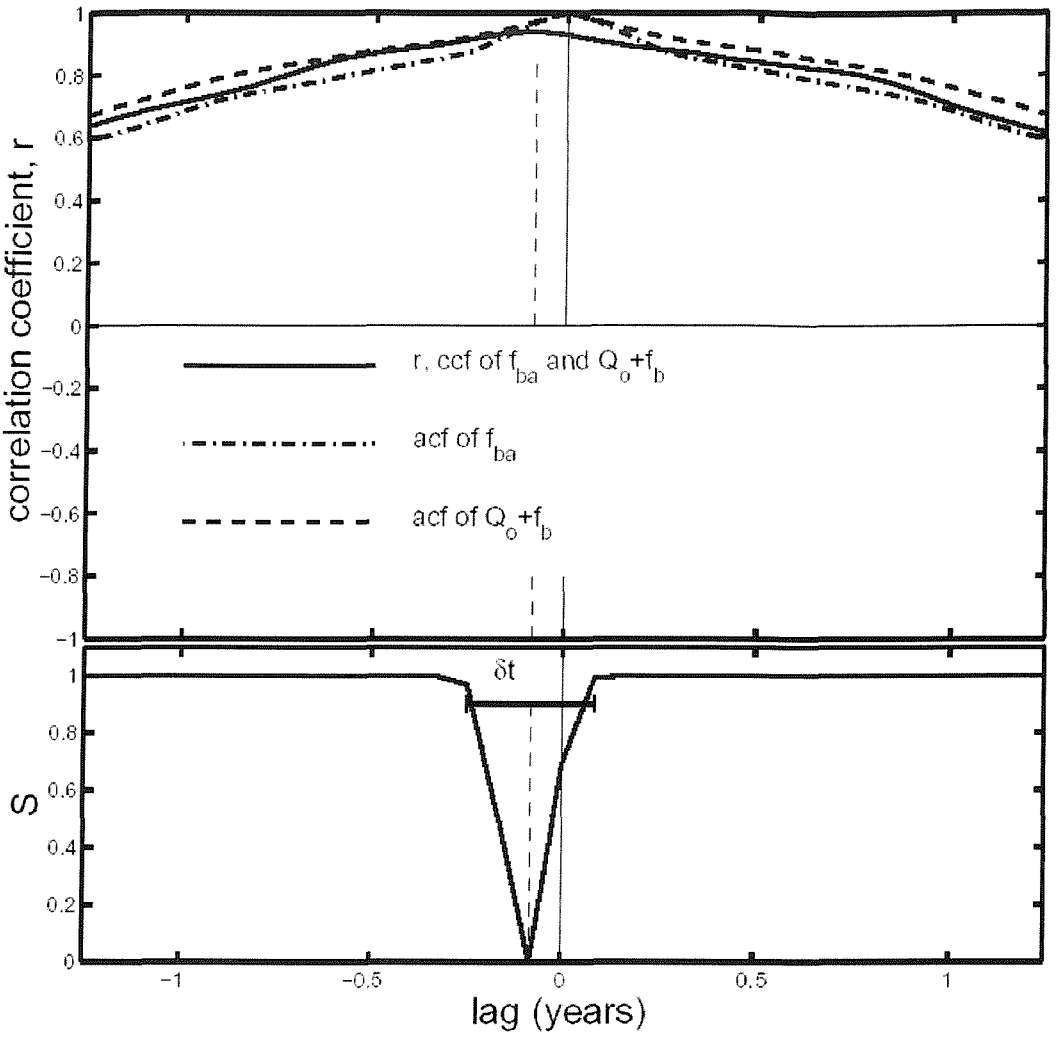


Figure 7.22. Correlogram of the reconstructed active-region facular brightening f_{ba} (from sunspot group data) and the observed brightening, $Q_o + f_b = I_{TS} + P_{SI}$. Three-month running means of f_{ba} have been used. The top panel shows correlation coefficients as a function of lag (defined as positive if f_{ba} leads $Q_o + f_b$): (solid line) the cross correlation between f_{ba} and $(Q_o + f_b)$; (dot-dash line) the autocorrelation of f_{ba}; (dashed line) the autocorrelation of $(Q_o + f_b)$. The peak cross correlation is at a lag of -1 month (vertical dashed line). The lower panel shows the significance S (evaluated using the Fisher-Z test) of the difference between the correlation at a general lag and the peak value. The error bar shows where the correlation is now significantly different from the peak value and defines the lag to be -1±2 months at the 90% significance level (S = 0.9).

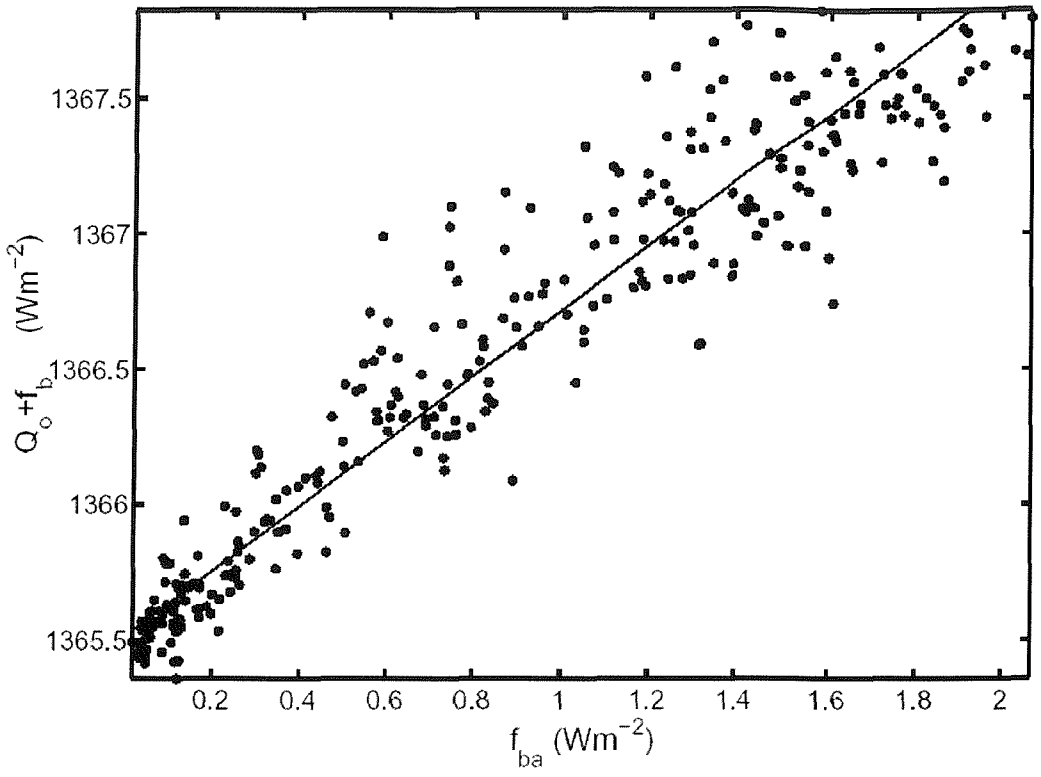


Figure 7.23. Scatter plot of three-month running means of the modelled active-region facular brightening f_{ba} against the observed brightening, $Q_o + f_b = I_{TS} + P_{SI}$ (see figure D). The correlation coefficient is $r = 0.953$, which is significant at the 98.0% level and explains $r^2 = 91\%$ of the variation. The solid line is the best linear regression fit ($Q_o + f_b$) = $s_a f_{ba} + c_a$. where $s_a = 1.20 \pm 0.06$ and $c_a = 1365.5 \pm 0.1$. The best fit is obtained if the f_{ba} data (derived from the sunspot group data) is lagged by one month (see figure 7.22).

The correlation coefficient is $r = 0.953$, which is significant at the 98% level and explains $r^2 = 91\%$ of the variation. This peak correlation is obtained with the f_{ba} data sequence (derived from the sunspot group data) lagged by one month. This lag accommodates the fact that the faculae in active regions persist somewhat after the sunspots, as noted in figure 7.1.

Because using method 2 gives us an estimate of active region brightening that is fully independent of the observed I_{TS} and P_{SI} we can use the results to deduce the brightening by network faculae. Equations (7.11) and (7.12) yields:

$$\Delta f_{bn} = I_{TS} + P_{SI} - (Q_o + f_{bno}) - f_{ba} = I_{TS} + P_{SI} - Q_{min} - f_{ba} \qquad (7.26)$$

where Q_{min} = 1365.5 is the modern-day solar minimum irradiance, which is the minimum monthly mean of (I_{TS} + P_{SI}) during the period 1978-2003 for which TSI and PSI observations are available.

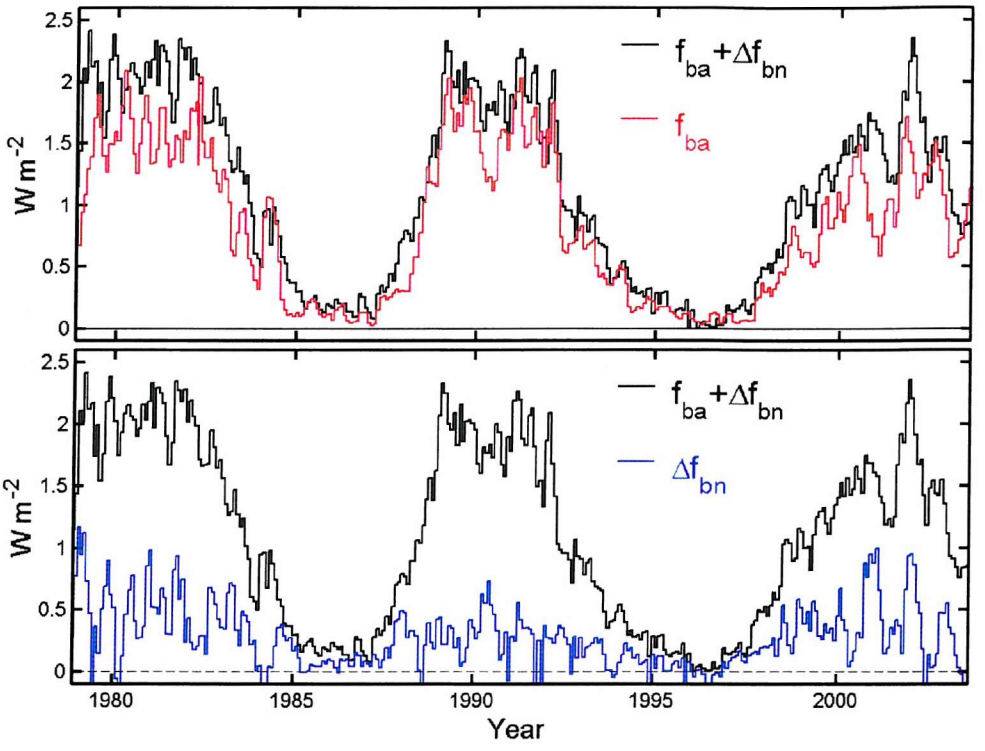


Figure 7.24. In both panels, the black histogram gives the observed variation of (I_{TS} + P_{SI} - Q_{min}) = (Δf_{bn} + f_{ba}). In addition, the top panel shows in red the f_{ba} variation, the lower panel also shows Δf_{bn} in blue.

The top panel of figure 7.24 shows, in black, the observed variation of (I_{TS} + P_{SI} - Q_{min}) = (Δf_{bn} + f_{ba}) along with the f_{ba} variation (three-point running means, delayed by the best-fit lag of 1 month) derived from the sunspot group data (in red). The lower panel of figure 7.24 again shows in black the observed (Δf_{bn} + f_{ba}) along with Δf_{bn} (in blue), derived by subtracting the f_{ba} variation. Figure 7.24 shows several peaks in the network facular brightening that follow on from peaks in the active region brightening. These are clearly faculae left over from active regions and hence are ephemeral flux rather than network. This effect is particularly pronounced for the most recent solar cycle (23), where three large peaks in Δf_{bn} can be seen in the figure following peaks in f_{ba}. The effect is also present in cycles 21 and 22, but is much less

pronounced. In several places, decreases in Δf_{bn} are simultaneous with peaks in f_{ba}, implying that f_{ba} has been overestimated.

The best-fit of f_{ba} to $(I_{TS} + P_{SI})$ has a slope of s = 1.20, which shows that the solar cycle variation in network facular brightening is 20% of that by faculae in active regions. Thus 17% of the total brightening over recent solar cycles is due to the network (and ephemeral flux) and 83% is due to active region faculae. Recent work by *Walton et al.* [2003] comes to a similar conclusion.

From the above, the total solar cycle variation in facular brightening is $f_{ba}/0.835 = 1.20 f_{ba}$. This is almost the same as $s_a f_{ba}$ derived from the regression fit in figure 7.23. Similarly, to within the measurement uncertainty, $c_a = Q_{min}$. Therefore, the regression fit shown in figure 7.23 effectively allows for the solar cycle variation of network faculae. Thus:

$$\Delta f_{bn} + f_{ba} = (1.20 \pm 0.06) f_{ba} + (Q_o + f_{bno}) \qquad (7.27)$$

where $(Q_o + f_{bno}) = 1365.5 \pm 0.1$.

Thus we can use the sunspot group data to predict the solar cycle variation in active region and network faculae right back to the start of the Greenwich data in 1874. The reconstructed variation of the facular brightening from the composite sunspot group data for 1874-2001 is shown in figure 7.25. Figures 5.11 – 5.14 show that the derived f_{ba} variation is homogeneous and not subject to any "glitches" because of the complex intercalibration of the Greenwich and Mt. Wilson sunspot data.

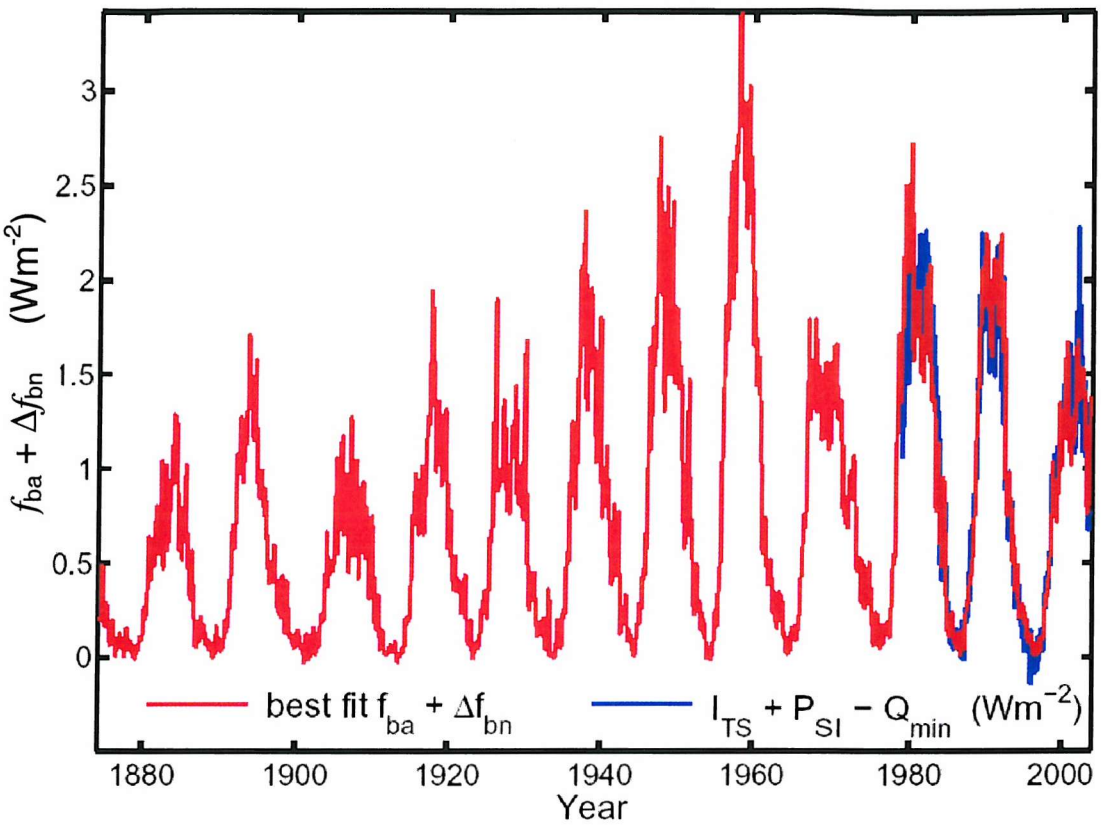


Figure 7.25. Reconstructed facular brightening ($f_{ba} + \Delta f_{bn}$, in red) derived from the f_{ba} from Greenwich/Mount Wilson sunspot group data for 1874-2001, and fitted to recent observations (in blue).

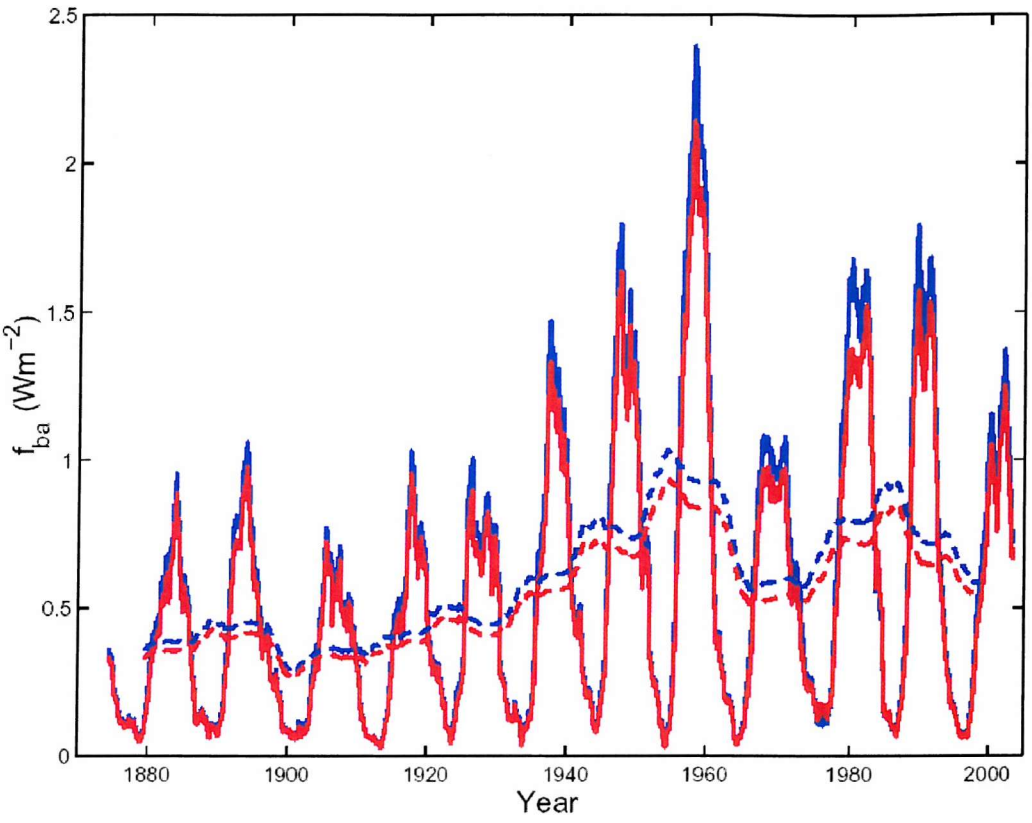


Figure 7.26. Annual mean of reconstructed active-region facular brightening f_{ba}, for (blue) active regions placed at the observed latitude and μ and (red) active regions placed at the solar equator and the observed solar longitude. The dashed lines are the corresponding 11-year running means. Values for the real latitudes are on average 9% higher than for zero latitude.

Figure 7.26, analyses the origins of the long-term drift and the relative role of active region area and of active region latitude. As discussed in chapter 3, average active region latitudes have risen over the past 100 years, as has the latitudinal spread of the active regions. As a result the average values of μ have fallen for active regions.

For the blue line in Figure 7.26, the active regions were assigned their true μ value (i.e., the active region is placed at the correct heliographic latitude and longitude) and f_{ba} computed from the sunspot group data, as described above. However, for the red line, each group was assigned to zero heliographic latitude and the correct longitude. It can be seen that there is an effect due to the latitudes of the active regions, which was predicted in section 7.3, with the increasingly higher

latitudes in more recent years helping to increase f_{ba}. The dashed lines show the 11-year running means of the corresponding colour solid lines. On average, 90% of the rise in $(I_{TS})_{11}$ (which here is all due to the rise in active region facular brightening) between 1900 and 1955 is caused by the rise in the number and area of the active regions (as seen in the red line), the remaining 10% of the blue line is caused by the broadening and poleward expansion of the active regions. Thus this is a significant, but certainly not a dominant effect.

Such an effect may be proportionally more significant if weak flux tubes of the extended solar cycle [*Harvey*, 1997] are a significant part of the f_{bno}, as seen at sunspot minimum. As yet we have no information on how much these contribute, and no information of how much their latitude may have changed over the past 150 years. However, the dependence of their brightness on μ could be proportionally more important than for the active region faculae analysed in figure 7.26.

Equation (7.11) shows that the total solar irradiance I_{TS} is the sum of: the irradiance of the field-free Sun, Q_o; the solar minimum facular brightening, f_{bno}; the solar cycle variation of network facular brightening, Δf_{bn}; the active-region facular brightening; f_{ba}; and minus the photometric sunspot index, P_{SI}. The sunspot data can be used to reconstruct P_{SI} and, from the above, $(\Delta f_{bn} + f_{ba})$. By assuming that irradiance changes are all caused by surface magnetic field effects, we have also over estimated Q_o. Thus the one unknown is the past variation in f_{bno} and below we generate three TSI reconstructions based on three assumptions about the behaviour of f_{bno}.

7.7.1. TSI Reconstruction 1.

There has been considerable debate about the possible drift in solar minimum irradiance, $Q_{min} = Q_o + f_{bno}$ over the recent cycles (f_{bno} is the network facular brightening at sunspot minimum). As discussed in Chapter 2, the ACRIM composite by *Wilson and Mordinov* [2002] shows a slight rise in Q_{min}, whereas the PMOD composite does not [*Fröhlich*, 2003a]. *Foukal* [2002] recently inferred that there was

little change in network faculae since the nineteen-twenties from analysis of historic observations of faculae: however, the effects of improved resolution and sensitivity of the observations make this comparison difficult. However, we can gain one extreme limit of irradiance behaviour (here termed reconstruction 1) if we assume that the both the quiet Sun and the background network brightening has remained constant.

If we assume that the quiet sun distribution of radial magnetic flux has remained constant, and so was the same as has been observed over the last half solar cycle by SoHO (the blue line in figure 7.4), we can use equations (7.1) and (7.2) to compute the distributions of radial field values for as given sunspot group area A_G. We here present the spectra of magnetic field values that this yields, by plotting the total flux threading the full photospheric surface in 1G bins (termed F_1) as a function of $|B/\mu|$ and date. Monthly spectra for recent cycles (when we also have TSI data) are shown in figure 7.27. The solar cycles can be clearly seen as a shift in the flux to higher values and an increase in the total emerged flux. Considerable structure is seen at this temporal resolution.

Figure 7.28 shows annual means in the same format for all the Greenwich/Mt. Wilson sunspot composite data, extending back to 1874. The effect of the lower emergence rates and lower solar activity 100 years ago is evident. These distributions of emerged flux could be used as model inputs to, or could be matched to predictions by, models of solar flux emergence and evolutions [*MacKay et al.*, 2002; *Mackay and Lockwood*, 2002; *Wang et al.*, 2002, *Lean et al.*, 2002]

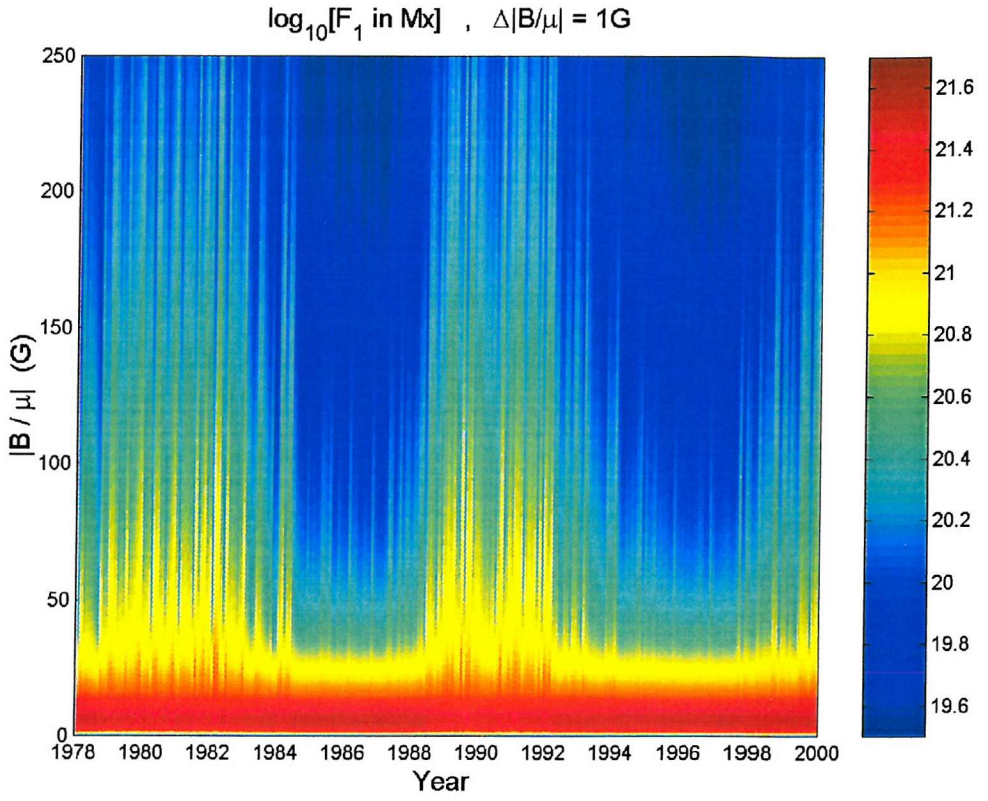


Figure 7.27. Distributions of radial surface flux, as a function of year and radial field per pixel |B/μ|. These distributions are monthly means derived from monthly means of the sunspot group area data from 1978 to 2000.

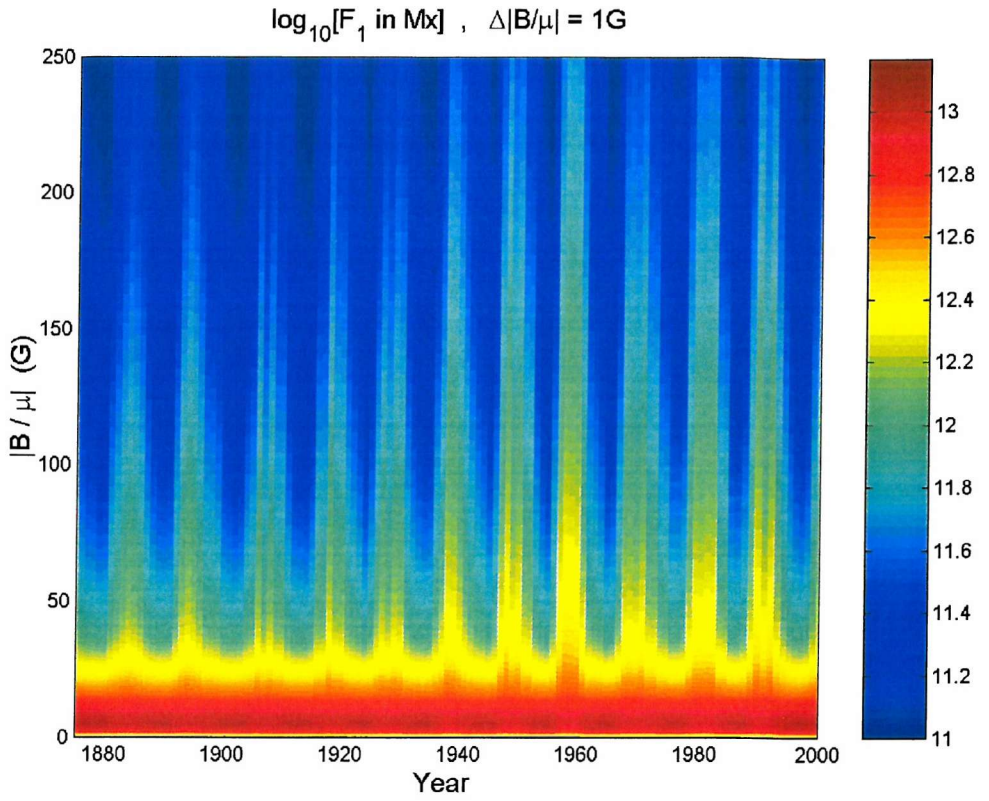


Figure 7.28. Distributions of radial surface flux, as a function of year and radial field per pixel $|B/\mu|$. These distributions are annual means derived from annual means of the sunspot group area data from 1874 to 2000 for reconstruction1.

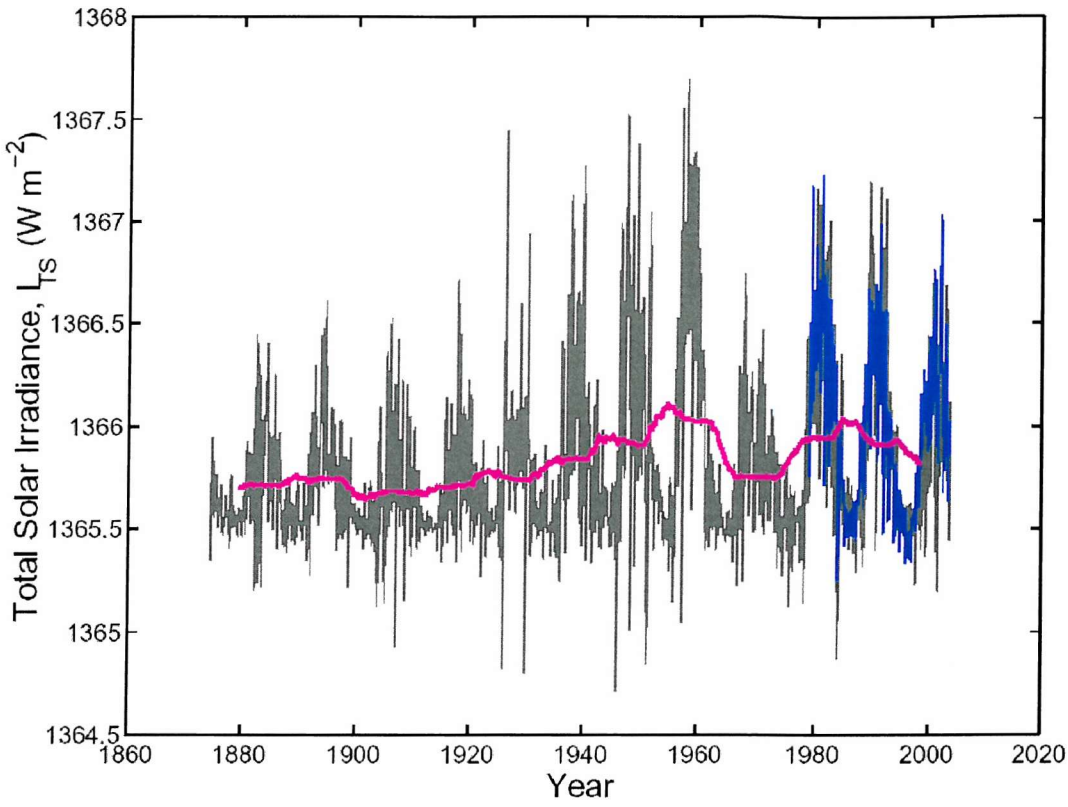


Figure 7.29. Reconstruction 1 of the TSI, I_{TS}, based on the assumption that the quiet sun irradiance Q_o and the solar-minimum network facular brightening, f_{bno} have both remained constant over the past 150 years. The grey line gives the monthly values and the blue line the observed values from the PMOD composite. The mauve line is an 11-year running mean.

Assuming that both Q_o and f_{bno} have remained constant, we obtain the irradiance reconstruction shown in grey in figure 7.29. The blue line are the I_{TS} observations and the mauve line is an 11-year running mean $(I_{TS})_{11}$ which reveals the average upward trend only because of the rise in the solar cycle amplitude of facular brightening, there being no change in the background network facular brightening or in the quiet Sun in this case.

7.7.2. TSI Reconstruction 2.

Additional long-term changes could arise from a drift in the solar-minimum network facular brightening, f_{bno}. To analyse this effect we change the width of the quiet Sun distribution. In this second reconstruction we define a second limit of the behaviour by assuming that there is no surface flux in the Maunder minimum.

To model the effect of variations in the quiet Sun, we assume that the distribution of radial field values away from active regions always has the same shape as in modern times, and as observed by the SoHO satellite. The distribution is then varied in width and then renormalized so that the number of pixels on the disk of the quiet Sun is constant. The results are shown in figures 7.30 to 7.33

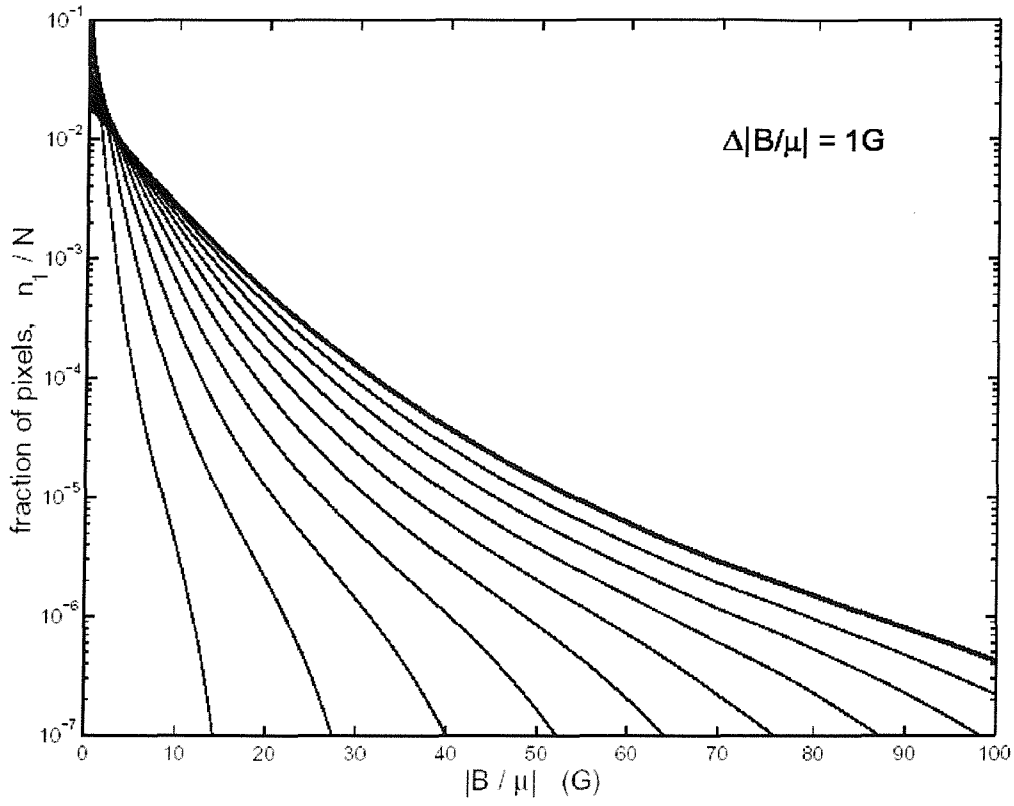


Figure 7.30. Model distributions of radial field values for the quiet Sun. The heavier line is the modern day distribution and the others are generated by reducing the width of the distribution and re-normalising to give a constant number of total pixels. The plot gives the fraction of pixels in 1G bins, n_1/N.

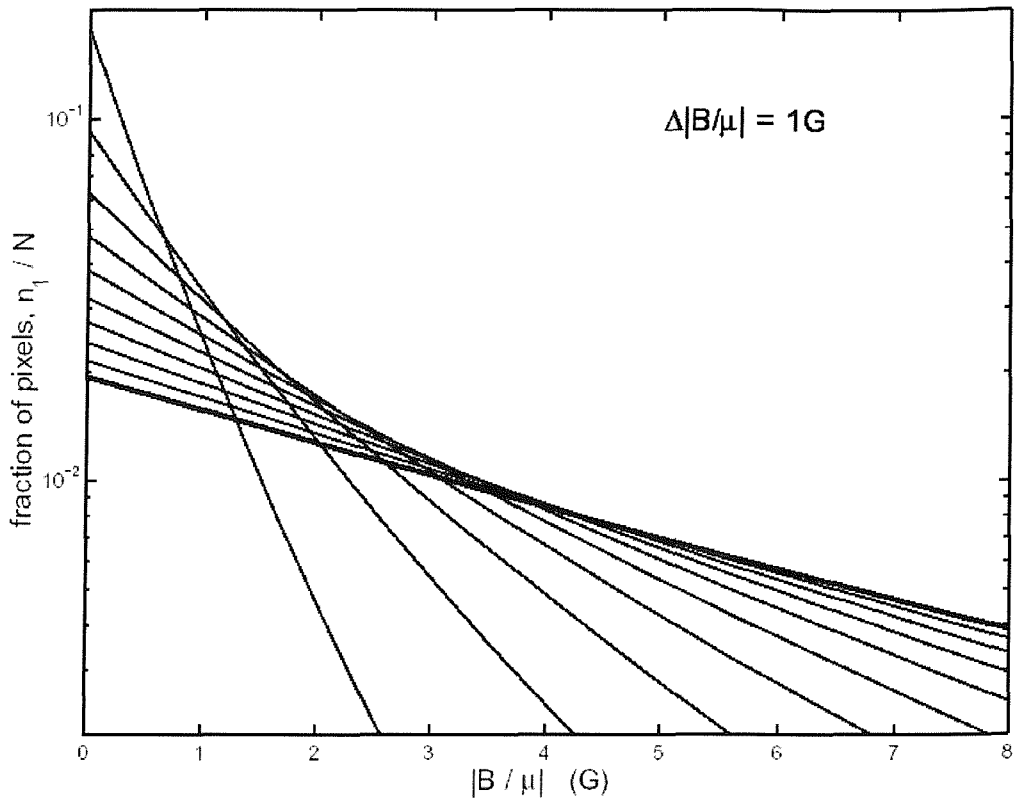


Figure 7.31. Detail of figure 7.30 at low radial field values, to show the increase in near-zero fields when the width of the distribution is reduced. The limit of zero width yield a delta function at $|B/\mu| = 0$.

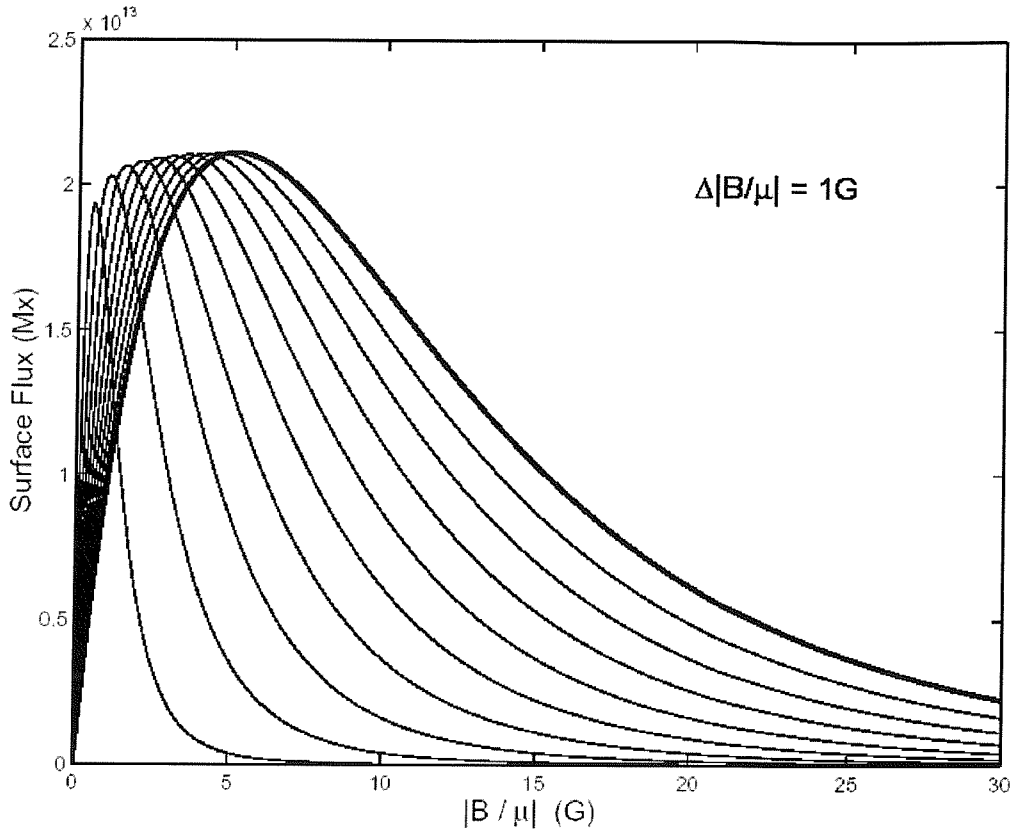


Figure 7.32. The total surface flux in 1G bins, F_1, as a function of $|B/\mu|$ for the quiet Sun distributions shown in figures 7.30 and 7.31.

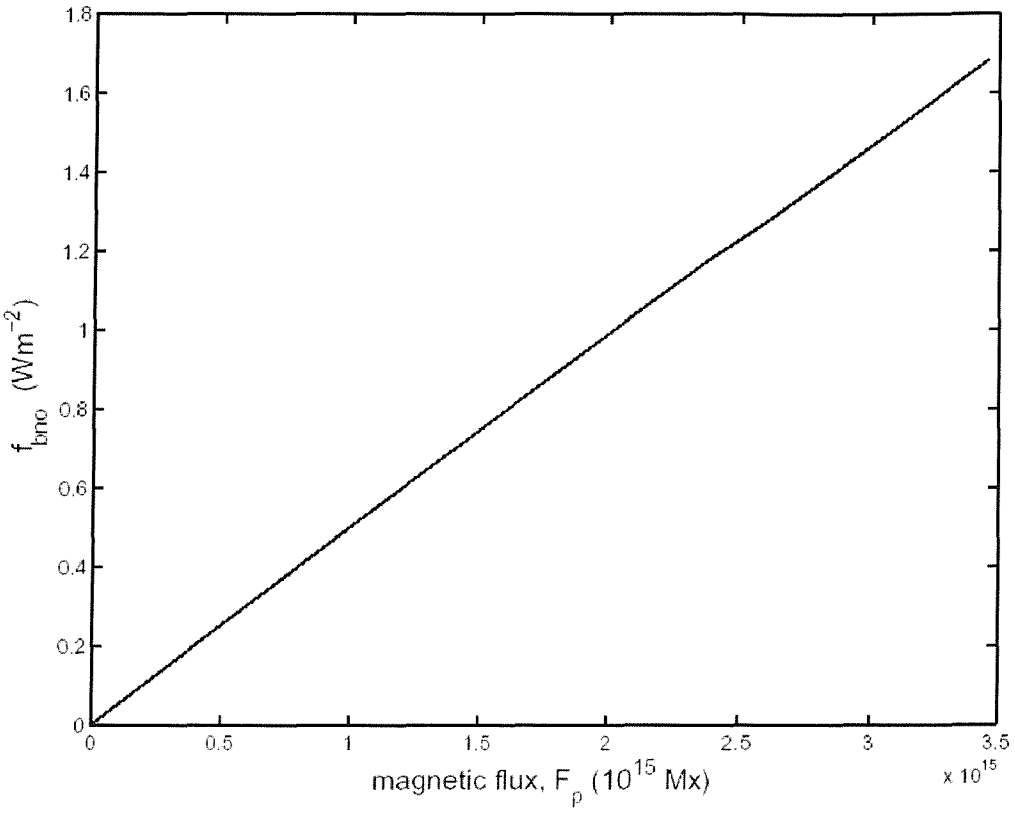


Figure 7.33. The disc-integrated facular brightening, f_{bno} (computed), as a function of the total surface magnetic flux F_p (at $|B/\mu|$ below 1000G) for the quiet Sun distributions, shown in figures 7.30 and 7.31. Because F_1 is for 1G bins of radial field, $F_p = \Sigma F_1$, where the sum is here carried out for ($0 \leq |B/\mu| \leq 1000G$).

Figure 7.33 shows a linear relationship of the quiet network (background) facular brightening with total flux. Such a linear relationship should have an analytic explanation.

The average contrast is:

$$\langle C \rangle = \left(\frac{1}{N} \right) \int c(B_r) n(B_r) dB_r \qquad (7.28)$$

where B_r is the radial field (= $|B/\mu|$) and the integration is carried out over the range used here ($0 \leq B_r \leq 1000G$). If the quiet Sun field distribution is uniform across the disc, integrating the *Ortiz* equation for contrast over all μ yields a full-disk contrast:

$$c(B_r) = \alpha B_r + \beta B_r^{\,2} + \gamma B_r^{\,3} \qquad (7.29)$$

where $\alpha = 2.32 \times 10^{-4}$, $\beta = -6.65 \times 10^{-7}$, $\gamma = 4.71 \times 10^{-10}$. From equation (7.28) and (7.29):

$$\langle C\rangle = \left(\frac{\alpha}{N}\right)\int B_r n(B_r)dB_r + \left(\frac{\beta}{N}\right)\int B_r^2 n(B_r)dB_r + \left(\frac{\gamma}{N}\right)\int B_r^3 n(B_r)dB_r$$

$$\langle C\rangle = \alpha\langle B_r\rangle + \beta\langle B_r^2\rangle + \gamma\langle B_r^3\rangle$$

$$\frac{\langle C\rangle}{\langle B_r\rangle} = \alpha\left[1 + \frac{(\beta\langle B_r^2\rangle)}{(\alpha\langle B_r\rangle)} + \frac{(\gamma\langle B_r^3\rangle)}{(\alpha\langle B_r\rangle)}\right] \qquad (7.30)$$

The second and third terms in (7.30) are given in Table 7.2 for the distributions given in figures 7.30 and 7.31, from the present-day quiet Sun distribution (w = 1) to the distribution that has a tenth of the present-day distribution (w = 0.1).

Table 7.2. Analysis of effect of width of quiet-Sun distributions.

w	$(\beta< B_r^2 >) / (\alpha< B_r >)$	$(\gamma< B_r^3 >) / (\alpha< B_r >)$	F_p (Mx)	f_{bno} (Wm^{-2})	f_{bno} / F_p ($Wm^{-2}Mx^{-1}$)	$< \lvert B/\mu \rvert >$ (G)
1	-3.9×10^{-2}	6.8×10^{-4}	3.46×10^{15}	1.682	4.87×10^{-16}	5.73
0.9	-3.5×10^{-2}	5.5×10^{-4}	3.11×10^{15}	1.519	4.89×10^{-16}	5.16
0.8	-3.1×10^{-2}	4.3×10^{-4}	2.76×10^{15}	1.353	4.90×10^{-16}	4.58
0.7	-2.7×10^{-2}	3.3×10^{-4}	2.41×10^{15}	1.187	4.93×10^{-16}	4.00
0.6	-2.3×10^{-2}	2.4×10^{-4}	2.06×10^{15}	1.019	4.95×10^{-16}	3.42
0.5	-1.9×10^{-2}	1.7×10^{-4}	1.71×10^{15}	0.850	4.97×10^{-16}	2.84
0.4	-1.5×10^{-2}	1.1×10^{-4}	1.36×10^{15}	0.680	4.99×10^{-16}	2.26
0.3	-1.1×10^{-2}	6.1×10^{-5}	1.01×10^{15}	0.507	5.01×10^{-16}	1.68
0.2	-7.7×10^{-3}	2.7×10^{-5}	6.64×10^{14}	0.334	5.03×10^{-16}	1.10
0.1	-3.9×10^{-3}	6.8×10^{-6}	3.16×10^{14}	0.160	5.05×10^{-16}	0.52

The second and third columns of Table 7.2 shows that these terms are much smaller than the first term in the square bracket in (7.30) which is unity. Thus to an accuracy better than 5% we can make the approximation:

$$\frac{\langle C \rangle}{\left\langle \frac{\beta}{\mu} \right\rangle} \approx \alpha = 2.32 \times 10^{-4} \tag{7.32}$$

yielding the irradiance change per unit total photospheric flux:

$$\frac{f_{bno}}{F_p} \approx \frac{(Q_o \langle C \rangle)}{\left(4\pi R_s^2 \left\langle \left| \frac{B}{\mu} \right| \right\rangle \right)} \approx 5.1 \times 10^{-16} \text{ Wm}^{-2}\text{Mx}^{1} \tag{7.33}$$

This is very close to the values derived from figure 7.33, as given in the penultimate column in table 7.2 (to within the predicted error of 5%). The last column gives the $<|B/\mu|>$ corresponding to the $F_p \approx$ value to allow easy comparison with other studies.

Thus to within a few percent, $<C>/<B_r>$ is independent of the distribution of B_r values. In theory, brightening changes could occur at constant total surface flux if the shape of the radial distribution also changes; however, in practice Δf_{bno} is proportional to ΔF_p to within a few percent. Thus quiet sun irradiance changes are almost all associated with changes in the total photospheric field.

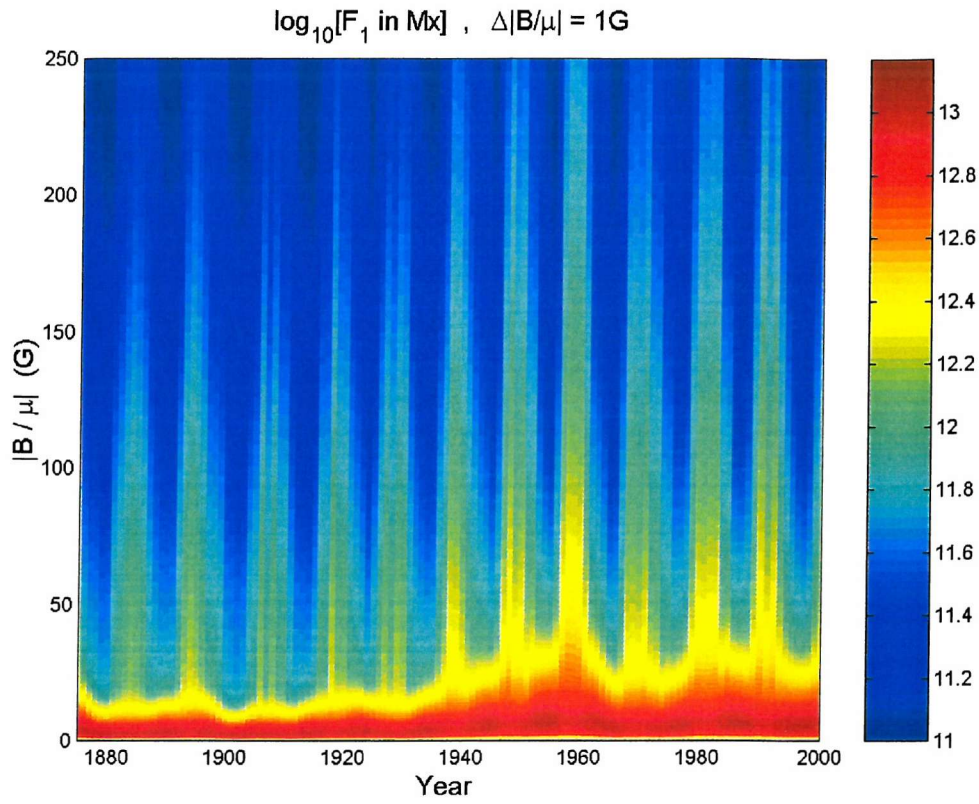


Figure 7.34. Distributions of radial surface flux, as a function of year and radial field per pixel $|B/\mu|$ for reconstruction 2. These distributions are annual means derived from annual means of the sunspot group area data from 1874 to 2000.

Lean et al. [1995] and *Lean* [2000a] use the 11-year running mean of sunspot number to give the variation of the background irradiance. We here assume that the variation of the background flux is proportional to the 11-year mean of the sunspot number, R_{11}, and hence are making essentially the same assumption as *Lean et al.* [1995] and *Lean* [2000a]. Table 7.3 gives the average surface fluxes implied for the Maunder minimum and for the 1901 solar minimum along with the changes between these dates and the present day values. Reconstruction 2 requires that $<|B/\mu|>$ was zero in the Maunder minimum, when R_{11} falls to zero, whereas reconstruction 3 (see below) requires that it was half present-day values.

Table 7.3. Average fields associated with the reconstructions

reconstruction	Maunder minimum		1901 solar minimum	
	$< \|B/\mu\| >$ (G)	$\Delta< \|B/\mu\| >$ (G) **	$<\|B/\mu\| >$ (G)	$\Delta<\|B/\mu\| >$ (G) **
1	5.73	0	5.73	0
2	0	5.73	2.15	3.58
3	2.865	2.865	3.94	1.79

** *Change with respect to present day values*

Figure 3.20 shows the results of simulations by *Lean et al* [2002], of the evolution of emerged solar flux under the influence of diffusion, meridional flow and differential rotation. The input emerged source flux has been increased linearly, roughly as inferred from sunspot numbers for much of the last century. The open flux rises as a consequence, but the rise in the surface field at solar minimum is much smaller. In fact, this simulation gives a rise of just 2 G in $<|B/\mu| >$, and this is smaller than the change of 3.58G required by reconstruction 2. (However, we note that it is consistent with the rise of just 1.79G required by reconstruction 3.)

Note, however, that the *Lean et al.* simulation, as shown in panel c of figure 3.20, only has flux emerging when sunspots are present. Thus "hidden flux" emerging in an extended solar cycle (with varying overlap) could increase the total emerged flux and its long-term variation and thus have the same effect on TSI.

Figure 7.34 shows the magnetic field for reconstruction 2, in the same format as 7.27 and 7.28 and figure 7.35 shows the corresponding TSI variation in the same format as figure 7.29.

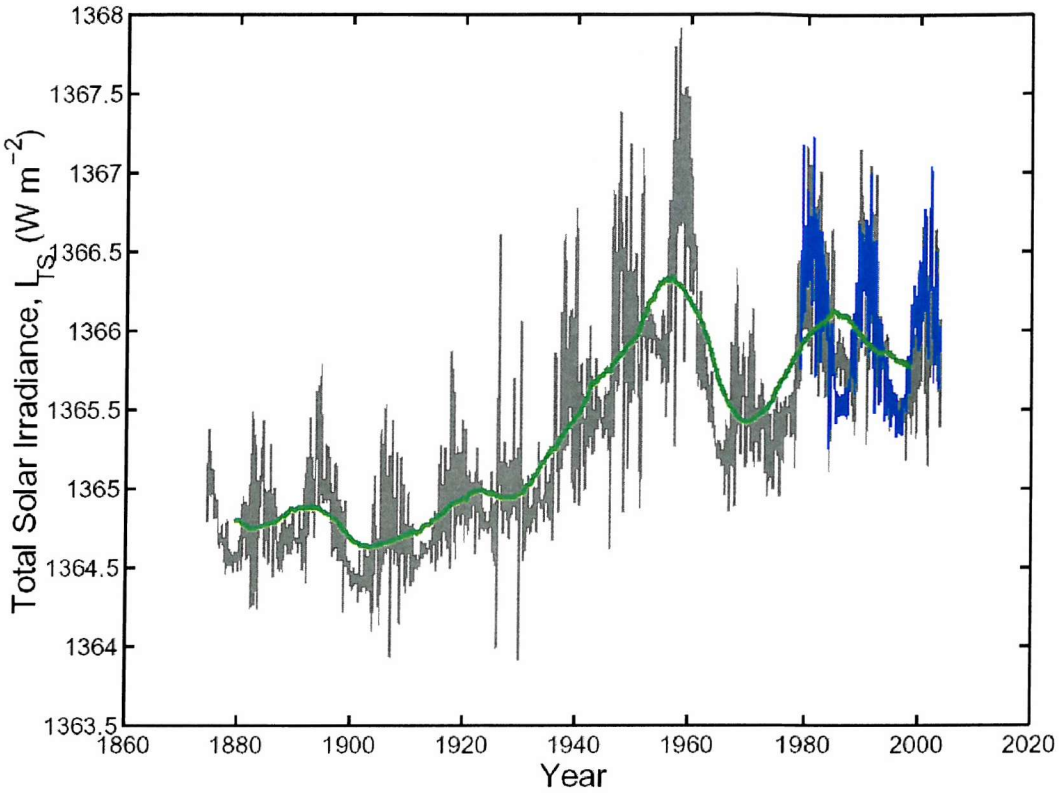


Figure 7.35. TSI reconstruction 2, based upon the assumption that surface magnetic flux during the Maunder minimum fell to zero. The grey line gives the monthly means, the blue lines the observed values of TSI, and the green line is the 11-year running mean.

Figures 7.29 and 7.35 are two extreme limits of irradiance variations, corresponding to no drift in the background surface flux and facular brightening and to all such flux and brightening being absent in the Maunder minimum. I next look at the variation in TSI if some residual surface flux continued through the Maunder minimum.

7.7.3. TSI Reconstruction 3.

It is unlikely that all surface magnetic activity ceased during the Maunder minimum, as assumed in reconstruction 2. There is clear evidence from ^{10}Be records (used as proxies for cosmic ray propagation, see chapter 6), sporadic sunspot observations and climate indicators that there was still some small-scale solar activity through this period. As has been shown (chapter 4), sunspots and variations cosmic rays (chapter 6) are features associated solar magnetic flux, and if the sun had been

truly magnetically inactive through the Maunder minimum it would be expected that sunspot activity would cease entirely, and there would be none or very little variation in cosmic rays. Since this was not the case, it is reasonable to assume that the sun was still magnetically active through this period and so $|B/\mu|$ would not have reached 0, thus TSI reconstruction 2 becomes a less attractive option.

It is also then difficult to take the opposite view point that there has been no change in solar activity since the Maunder minimum. The majority of surface flux emerges in active regions and then disperses and moves under supergranular diffusion and larger-scale flows. Thus it is reasonable to expect the flux outside active regions to show some variation that mirrors the active region flux (with a delay of order 1 year)

To this end in TSI reconstruction 3, I have taken an average of the two limits of behaviour that have been discussed. I have assumed that the $|B/\mu|$ magnetic field during the Maunder minimum would have been half the present day values, leading to the distributions that can be seen in figure 7.36. I then assumed (see table 7.3) that values of $|B/\mu|$ had doubled to its present day values. The variation in Q_o is then calculated as for reconstruction 2, creating the TSI reconstruction that can be seen in figure 7.41. As can be seen the TSI increase over the past 100-year was not as sharp as reconstruction 2, with an overall increase of around 1 Wm^{-2} being found over the past 100 years. This is compared with the solar cycle variation that is known to occur, which is also around 1 Wm^{-2} (0.1%), and it can be said that the present day minima are around 0.1% larger than turn of the century values.

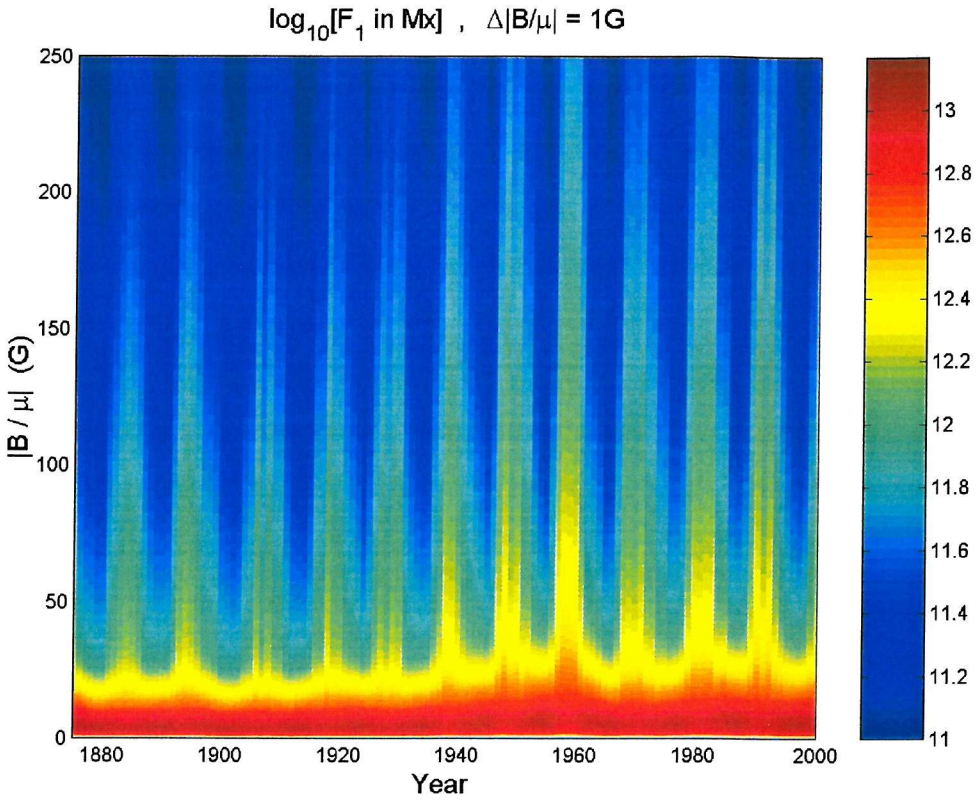


Figure 7.36. Distributions of radial surface flux, as a function of year and radial field per pixel $|B/\mu|$ for reconstruction 3. These distributions are annual means derived from annual means of the sunspot group area data from 1874 to 2000.

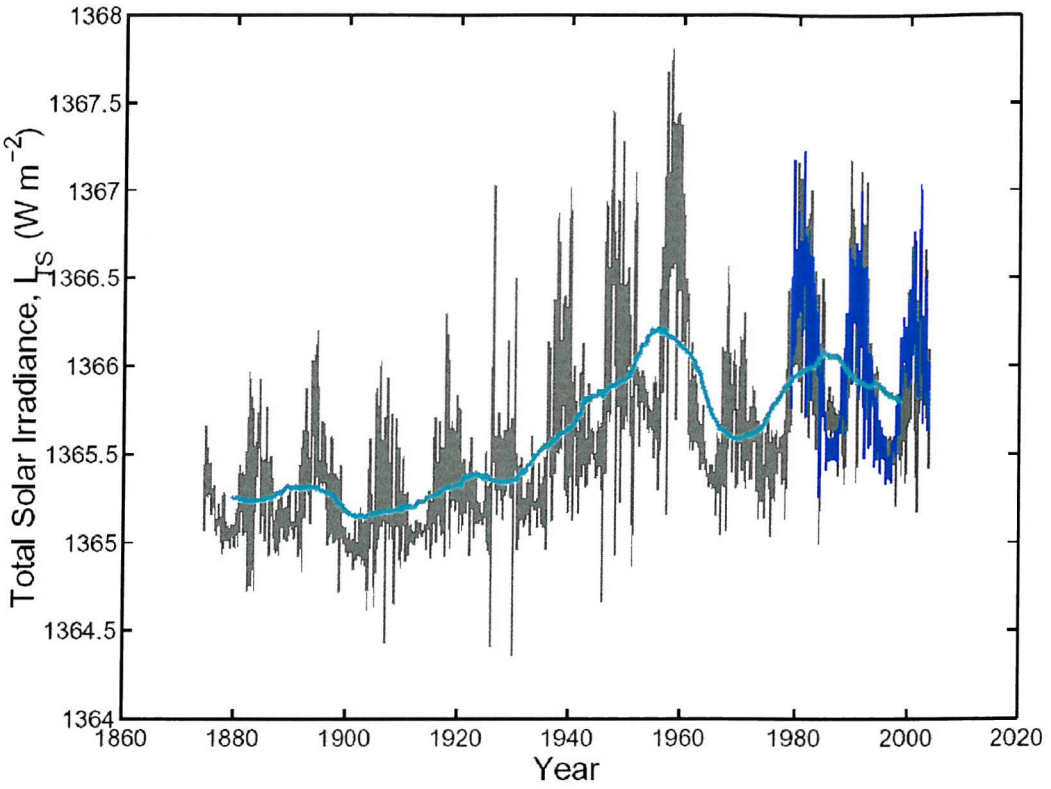


Figure 7.37. TSI reconstruction 3, based upon the assumption that surface magnetic flux during the Maunder was half the present day values. The grey line gives the monthly means, the blue lines the observed values of TSI, and the cyan line is the 11-year running mean.

7. 8. Implications of the TSI reconstructions.

I have defined TSI reconstructions 1-3 for which $\Delta I_{TS}/\delta I_{TS}$ is 0, $\approx$ 1.8 and $\approx$ 0.9, respectively. Reconstructions 1 and 2 reveal the extreme limits of the behaviour, either that the quiet Sun has shown no variation in irradiance since the Maunder minimum, or that the quiet Sun has varied by the maximum amount allowed by the *Ortiz et al.* [2003] contrasts since the Maunder minimum. Both of these show a long-term increase in irradiance over the past 100 years (when looking at the eleven year running means), although the increase when the quiet Sun has remained constant is small. Since the climatic system is believed to smooth out decadal irradiance variations, due to the heat capacity of the oceans, it is the smoothed long-term increases that are of particular interest to climate studies.

Reconstruction 1 shows that even if the quiet sun does remain constant, TSI can still increase long-term, just due to the increasing TSI variation over the individual solar cycles. Thus it is not necessary in climate studies for the background solar irradiance to have changed to cause the increases in global temperature, simply that the variation in TSI has increased over successive solar cycles. However, this calls for the global climate system to be more sensitive to small changes in irradiances than deduced from detection – attribution techniques using a TSI reconstruction that has a larger long-term drift. For example, *Stott et al.* [2000] use the HAD3CM global coupled model and show that the *Lean et al.* [1995] reconstruction calls for an amplification "Beta" factor of 2.5 - 3 to match global temperature distributions and their century-scale changes. For a reconstruction with a much smaller drift (such as reconstruction 1) this factor would be larger (assuming the detection – attribution procedure could detect the smaller solar variation at all).

Reconstruction 2 and 3 show a constant ramping up of the solar magnetic field, which cause large long-term solar irradiance variations. Here the long-term increase is dominated by the increases in the background quiet sun, with the solar cycle variation playing a more minor part. Given that even reconstruction 2 has a smaller long-term drift than *Lean et al.,* [1995], the conclusion about the Beta factor discussed above would also apply to reconstructions 2 and 3.

Figures 7.38 and 7.39 show the three reconstructions outlined in the previous section plotted with TSI reconstructed using the open flux (as discussed in chapter 6). In addition $\langle TSI \rangle_{11}$, the eleven year running means of monthly TSI data is shown in black. As can be seen reconstruction 2 fits closely to the eleven-year running means of the TSI observations.

Figure 7.39 gives a close up of the period when the PMOD composite TSI data is available. Reconstruction 1 (mauve) is found to over estimate irradiance at solar maximum, and underestimate irradiance at minima, and reconstruction 2 (green) is found to do the reverse. Reconstruction 3 is found to replicate the observed drift in TSI variation extremely well, adding more weight to the argument that this reconstruction best replicates the long-term drift in TSI. The grey triangle in figure 7.41 shows the extent to which instrumental drift could cause the variations in TSI

shown in the black line, as can be seen the drift lies outside this drift, showing that the irradiance variation seen must be real. This grey area is based on 3 ppm yr^{-1} quoted by Fröhlich [2000], but for these long-term drifts this is pessimistic and a value nearer 1ppm yr^{-1} is probably more appropriate. Thus the downward trend shown in figure 7.43 (which is well matched by reconstruction 3) is thought to be real.

Figures 7.38 and 7.39 also show two TSI reconstructions based on the open solar flux, F_S, the relationship of which to TSI was discussed in chapter 6. the dashed grey line is 11 year running means of the *Lockwood and Stamper* [1999] reconstruction, i.e. it is based on the correlation between monthly or annual TSI and F_S values. This reconstruction assumes that this correlation holds on century as well as decadal timescales (see chapter 6). Figures 7.15 and 6.4 show that this is very similar to the *Lean* [2000a] TSI reconstruction.

The dark blue lines in figures 7.38 and 7.39 are obtained by fitting the 11-year running means of F_S, $\langle F_S \rangle_{11}$, to the corresponding means in TSI, $\langle I_{TS} \rangle_{11}$. Thus it is not a fit to the solar cycle variation but to the drift in TSI observed since observations began. Because we only have 30 years of TSI data the uncertainties in this fit are large. Despite the uncertainties in both the fitted open flux variations and in the TSI reconstruction, an interesting point emerges from 7.38. The best fit of the smoothed open flux variation agrees rather well with reconstruction 2 in both form and amplitude, but in fact all reconstructions agree rather well in form, whether they are based upon open flux or the sunspot group data composite. So figure 7.38 gives justification for the use of cosmogenic isotopes as a linear indicator of TSI change in paleoclimate studies. However, because of uncertainties about the hidden flux and the total surface flux in the Maunder minimum we do not know the scaling factor that should be applied to the open flux/cosmic ray/cosmogenic isotope data to give TSI.

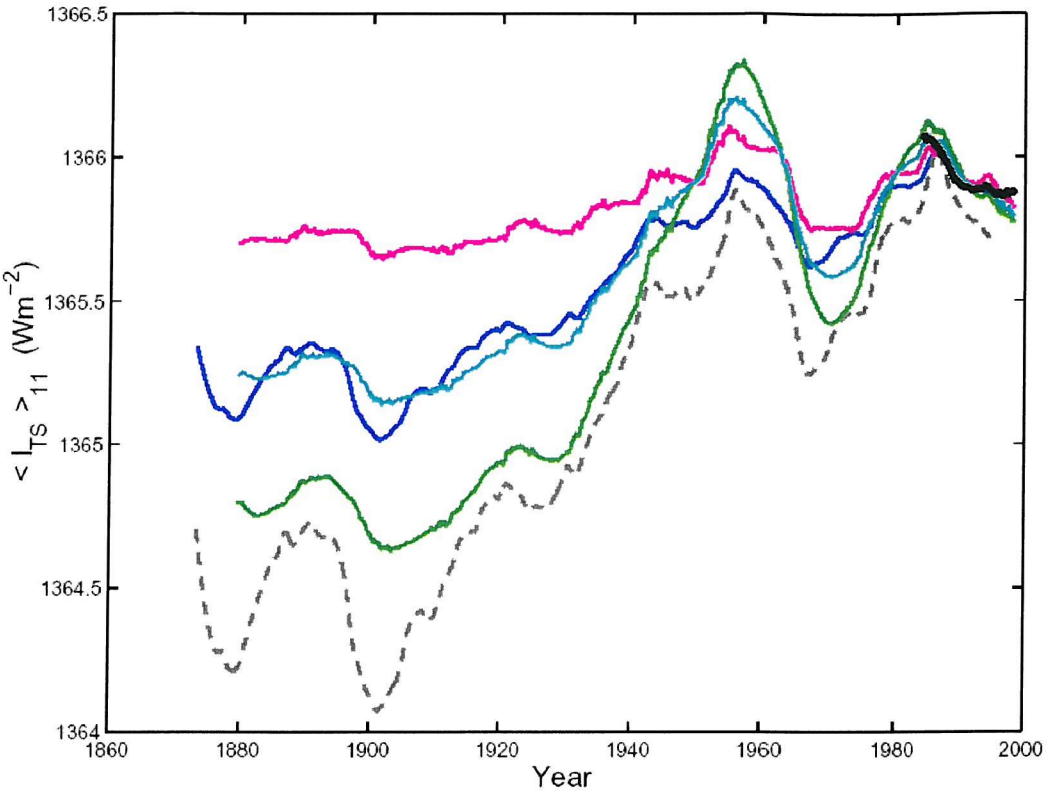


Figure 7.38. Eleven-year running means of the total solar irradiance, I_{TS}. (Mauve) from reconstruction 1, (green) from reconstruction 2 and (cyan) from reconstruction 3. The short, thick black line is from the PMOD composite of direct observations. The blue line shows the 11-year running means $\langle F_s \rangle_{11}$ of the open flux deduced from the aa index, F_s , which fitted to the 11-year running means of the PMOD composite of the observations. The grey-and-cyan dashed line shows 11-year running means of the irradiance from the fit of monthly F_s values and the monthly PMOD I_{TS} values.

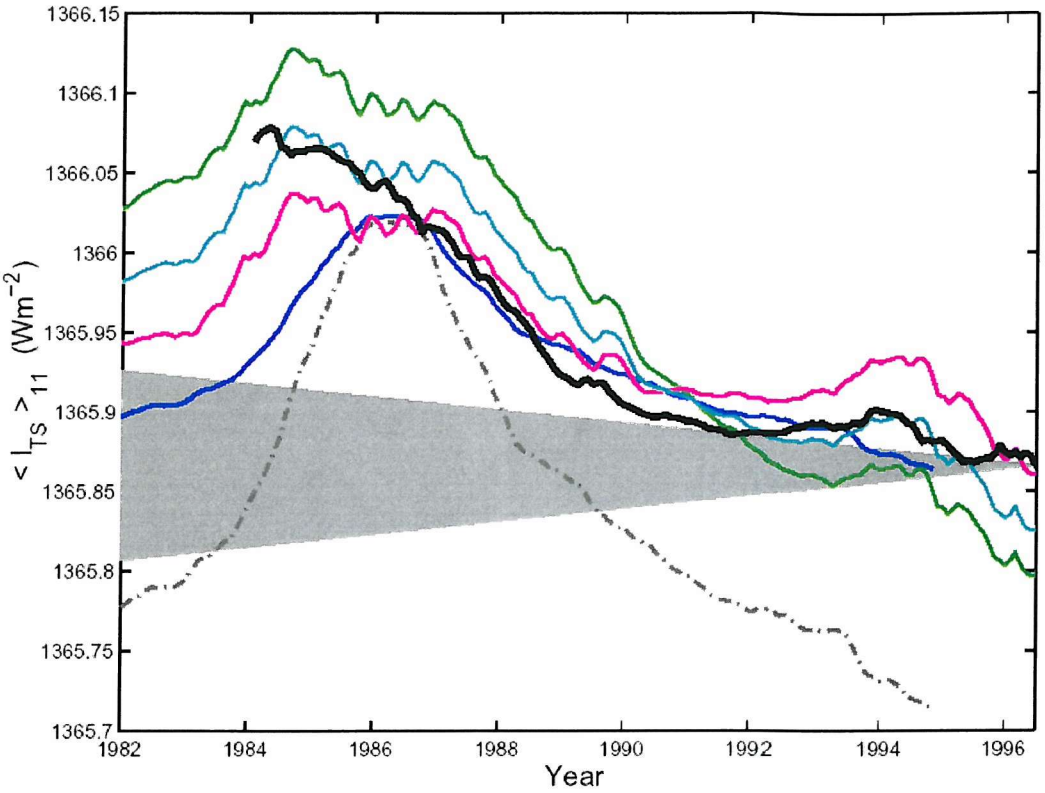


Figure 7.39. Close-up of figure 4.38 for 1982 to 1996. The grey area shows the extent of possible instrumental drift in the composite of observed I_{TS} values for of 3ppm per year.

7. 9. Chapter summary.

It has been found that the area and distribution of large flux tubes (sunspots) can be used to model the distribution and area variations of small flux tubes (i.e. faculae)and using the equations of *Ortiz et al.* [2002] we can predict the facular brightening that they cause.

Using averages of contrast variations with position, and the reconstruction of facular area from the sunspot group data, total facular brightening in active regions, f_{ba}, has been modelled with great success, and it found to match up closely to the observed variations in facular brightening (and quiet sun) variations as calculated in Chapter 6. It has been possible to estimate the variation in network faculae associated with the solar cycle, and the network faculae brightening that is part of the constant solar background. Thus we have been able to calculate the variation of the quiet sun

variation since 1978, and have been able to specify a minimum value of the quiet sun irradiance during the Maunder minimum as Q_0 = 1363.8 Wm^{-2}. This fully quiet sun irradiance is the value at which the sun is thought to be decrease to when there are no features associated with irradiance variations on the surface of the sun. This variation has been investigated with respect to a number of TSI reconstructions (including the reconstructions calculated in chapter 6) and nearly all reconstructions are found to drop below this minimum value, thus although not making these reconstructions incorrect, it does raise doubts about them, and reworking may be needed. However, the "hidden flux" effect in magnetogram pixels may mean that this estimate of Q_0 is too high, and I estimate that Q_0 could be as low as 1363.4 Wm^{-2}.

Using possible long-term variations and increases in surface magnetic flux $|B/\mu|$, we have created 3 reconstruction of TSI incorporating the variations of f_{ba}, calculated from sunspot group records, corresponding to no magnetic change, full magnetic change, and half magnetic change (in accordance to the *Lean* model) since the Maunder minimum. These reconstructions have all shown a long-term increase in irradiance since the turn of the century, with varying degrees of drift, and reconstruction 3 (which requires the surface field in the Maunder minimum to have half present day values) fits recent long-term trends in observed TSI very well.

Although unknowns and uncertainties prevent the calibration of open flux (and thus cosmogenic isotopes) in terms of TSI, all the reconstructions show a monotonic, almost linear relationship between TSI and open flux. Thus cosmogenic isotopes can be used as a proxy for TSI variations. (Note however, that there may be other effects, which also influence cosmogenic isotope abundances and others, which are relevant to the relationship between these abundances and climate changes).

Chapter 8: Conclusions.

8.1 Summary.

Chapter 3 of this thesis investigates relationships of surface magnetic features on the Sun to variations in the open solar magnetic field in the heliosphere. The reason for undertaking such a search is to investigate potential links between the open solar flux and total solar irradiance. Such a link has been suggested by correlations over recent solar cycles and has often been, implicitly or explicitly, assumed by paleoclimate studies employing cosmogenic isotopes.

The results confirm the long-term variation in annual sunspot latitudes first reported by *Pulkkinen* [1999], and also shows the same the result for monthly data. This long-term increase is found in both the Greenwich and Mt. Wilson data sets and the overlap periods between the two data sets are found to match extremely well, such that a continuous data set is created of mean sunspot latitudes from 1874 to the present day, for both the annual and monthly data sets. The period of greatest sunspot group latitude is found to occur in solar cycle 19, which corresponds with the highest known peak in a number of other solar variables, such as sunspot number and sunspot group area. The analysis of cosmogenic isotopes by *Usoskin et al.* [2003] suggests that that these peak values are the largest that have occurred in the past 1200 years.

The standard deviation in sunspot group latitudes is calculated and this is a measure of the area of the solar photosphere over which the sunspots can occur, I refer to this as the "sunspot spread". This spread is found to have increased in both annual and monthly data since 1874, again with the largest values being found in solar cycle 19. Interpretation of the spread data set is made difficult, especially for the annual data, when old cycle (low latitude) spots occurred in the same time bin as new cycle (high latitude) spots. A method to deal with this problem was developed, involving defining the artificial separator between the latitudinal distributions of new and old cycle spots, calculating these bands separately, then combining the two groups in each hemisphere together. These periods of overlap have been found to

increase with the activity of the Sun, although it is unclear whether old cycle spots are persisting longer, or new cycle spots are occurring earlier. Periods of overlap of smaller flux tubes in an "extended solar cycle" would be significant as they cause a variable solar minimum brightening which would contribute to long-term drifts in TSI.

Monthly and annual sunspot spread is found to correlate very well with the long-term variation in the coronal source flux, F_S as derived from the geomagnetic aa index by *Lockwood et al.* [1999a]. This long-term drift in open flux is now understood in terms of the observed variations of emergence rate in active regions being reproduced in numerical simulations by *Wang et al.* [2002], *Lean et al.* [2002] and Schrijver et al. [2002]. This correlation is found to be stronger for annual data than monthly data, but the significance is found to increase for monthly data. Sunspot spread has been used in conjunction with the equations of *Solanki et al.,* [2000], to recreate the variation in the emergence rate of magnetic flux at the solar surface, E, and this procedure is found to replicate the observed coronal source flux well. The link between coronal source flux and emergence rate depends on the loss rate and using sunspot spread to quantify emergence rate requires a loss time constant of 3.6 years, similar to that found by *Solanki et al.,* [2000], who used sunspot number to quantify E. This long time constant has been controversial as many scientists contend that open flux does not last this long. However, the Solanki et al. model is a simple continuity equation of all open flux, irrespective of where it is distributed on the solar surface. Thus it contains recently emerged open flux with footprints in the active regions and much older open flux, which has taken a year to migrate to the polar coronal holes. Some of this flux remains in the polar coronal hole for almost a whole solar cycle before it is destroyed by open flux of the opposite polarity emerging in the next solar cycle. On the other hand, some open flux is destroyed very quickly at the PIL of the BMR or at current sheets between BMRs. Thus averaging over all open flux a loss time constant of order 3 years may not be unreasonable. These results suggest that sunspot spread correlates better with the total emergence rate of flux through the solar surface than, for example, sunspot number or sunspot group area because if defines the area over which flux emergence can occur in both sunspots and faculae.

In chapter 4, I reviewed the theory of the effects of magnetic fields on solar luminosity and radius. These can be split into two separate effects. The first is surface effects, where magnetic fields cause variations in the visible solar output by modulating the emissivity of the surface. These effects are dominated by sunspots and faculae. The photometric effect of faculae and sunspots on the surface irradiance can be calculated for each feature individually using the mathematical models developed. Modelling of TSI variations from magnetogram data has been very successful if the magnetogram is broken down into just three components: quiet sun, sunspots and faculae [*Solanki and Fligge*, 1998]. Improvements are minor if we move to 4-component model which treats umbrae and penumbrae separately [*Kivrova et al.*, 2003] or a 5-component model, which treats network and active region faculae separately. Magnetic fields may also have effect deeper in the convection zone leading to sources and sinks and subsurface variations in heat transport towards the surface. These may cause "shadow effects" in the surface intensity. However the extraordinary success of the 3-, 4- and 5-component models of surface emissivity effects does not leave any room for shadow effects, at least on decadal timescales. The theory of *Spruit* [1982a,b, 1991] does not predict long-term shadow effects as extra heat is supplied from below to compensate for the sink or blockage. The bright rings around sunspots have not been specifically included because the depth at which spots cause this bright ring was believed to be too shallow from the mathematical theory developed by *Spruit* [1976]. The existence of some bright rings implies turbulent diffusion and convection are more effective at bringing heat flux to the surface than the *Spruit* theory predicts. One interesting possibility raised by the theory is that heat flux blocked by magnetic flux can be stored in the convection zone and released over the timescale of years (the diffusion timescale of the convection zone). Thus although irradiance is decreased due to the effects of sunspots, some of it could later be returned, increasing the quiet sun irradiance.

In chapter 6, monthly variations in the coronal source flux were calculated using the method of *Lockwood et al.* [1999a] and were found to correlate well with the PMOD composite of total solar irradiance, confirming the link found in the annual values found by *Lockwood and Stamper* [1999] and studied by Lockwood [2002a]. Using this correlation, I used the open flux as a proxy for TSI and thus I was able to reconstruct the long-term variation in TSI for the period 1868-2000. This

reconstruction is found to replicate the variations in TSI independently inferred by *Lean* [2000a] from a combination of sunspot number and it's eleven year running mean. This implies the correlation applies on century timescales as well as the decadal timescales on which it was first noted. Annual variations in the flux emergence rate, E, inferred from sunspot number are used reconstruct variations in F_S back to 1700 and thus using the same method applied to the monthly variation in F_S, TSI can be reconstructed back to 1702. From this reconstruction a value of 1364.1 Wm^{-2} is estimated for the TSI during the Maunder minimum, which is consistent with values inferred from stellar analogues and Sun-like stars in non-cyclic states

Cosmic rays and TSI are found to be strongly anti-correlated, and although it will involve the open solar flux, the physical link between these two parameters remains unclear. This anticorrelation is often implicitly or explicitly assumed in paleoclimate studies [e.g. *Bond et al.,* 2001]. Variations in the ^{10}Be isotope are used to reconstruct the long-term variation in TSI, back to 1424, and it can be seen that values during the past 3 solar cycles have been the highest in this interval. The smallest TSI values are found to occur during the Maunder minimum, for when I find a value of 1364.6 Wm^{-2}.

The photometric sunspot index, PSI, developed by *Fröhlich et al.*, for the period 1981-2000, has been shown to correlate strongly with sunspot area and monthly values of PSI have been reconstructed for the period 1874-2000, using the Greenwich and Mt.Wilson data sets. From this reconstruction it was shown that maximum values of PSI have increased steadily since 1900, with a maximum value being found in solar cycle 19. This reconstruction of PSI was continued in monthly form with the use of sunspot number back to 1702, and annual data back to 1700, due to the high correlation found between sunspot number and PSI.

Using the reconstructions of both PSI and TSI, the monthly variations in facular brightening and quiet sun was calculated on the simple basis that TSI was the product of quiet sun irradiance, the increase in irradiance due to faculae, and the darkening effects of sunspots. For recent years the reconstruction of $(Q + f_b)$ was found to correlate extremely well with known indicators of facular brightening, such as monthly facular area. Again this reconstruction was found to have a similar long-

term drift to that found in TSI. Annual ($Q + f_b$) values were calculated from the annual values of TSI and PSI and it was found that the values during the Maunder minimum were similar to that found in TSI, giving us a lower limit on the possible value of the quiet sun irradiance of 1364.1 Wm-2. The long–term (since 1800) drift ($Q + f_b$) was found to be too large to be readily attributed to network faculae alone, suggesting that the quiet Sun must be having some effect upon long-term drifts in both TSI and ($Q + f_b$), implying that shadow effects may well be needed to explain the variations on century timescales.

In chapter 7, I have used the equations of *Ortiz* to evaluate the variation in facular brightening by utilising the Greenwich sunspot group records, and the variation in the contrast of singular pixels in a SoHO MDI image. The contrast is found to vary with position on the solar surface (μ) and the field within the magnetogram B_{MDI}. Using the relationship between sunspots and active region faculae, positions and areas of faculae were determined and the contrast variation of each active region as it passed across the surface of the sun was calculated. This led to the brightening due to each individual sunspot group being calculated and a long-term variation in active region faculae being determined.

This then allowed me to directly calculate the quiet sun irradiance in the absence of surface magnetic effects, which I found to be 1363.8 Wm^{-2}, which then gives a lower limit value for the solar irradiance during the Maunder minimum, when it assumed that the sun was at it's lowest magnetic activity for the past 3 centuries. The only ways that lower values would have been possible is if shadow effects are an important factor or if the Ortiz et al. contrasts are too small because of the effect of cancelling "hidden" flux within many magnetogram pixels. A first-order correction for the hidden flux effect was made. From this I was able to investigate a number of models to see if they satisfied this limit of irradiance and found that nearly all exceeded this limit and so require non-surface effects in addition to the known surface effects. The closest model was that of *Lean* [2000a] which required the smallest non-surface effect.

Using the Greenwich data, I then developed 3 new and independent models of the long-term variations in TSI. These models vary depending upon the input of the

possible surface magnetic flux variations since the Maunder minimum. The greatest long-term drift in facular brightening is found if it assumed that Sun was essentially magnetically quiet. The smallest long-term change is found if it is assumed that the network faculae have not changed since the Maunder minimum. I believe that the best model is created using the compromise assumption that the network magnetic flux was half it's present day values during the Maunder minimum. These 3 reconstructions were compared to the variation in TSI calculated from the long-term variations in the coronal source flux, and the PMOD composite. The upper and lower limit variations of magnetic variations were found to over and under estimate the drift in decadal averages of the observed TSI over the last 2.5 solar cycles and the best-fit was obtained using the 'compromise' assumption about the Maunder minimum.

The key point about the reconstructions presented here is that they do not use a simple unexplained correlation (such as with open solar flux) nor do they rely on a drift amplitude derived for stellar analogues. Nevertheless, the drift in TSI on decadal timescales (for all 3 reconstructions) correlates well with the open solar flux variation found by *Lockwood et al* [1999a]. This gives a much needed justification for using cosmic rays and cosmogenic isotopes as proxy indicators of TSI in paleoclimate studies.

8.2. Future work.

This work highlights a number of areas that need to be researched in more detail. As has been seen, there are a number of parameters that correlate well, but the physical links between them are unclear. The evidence is ambiguous as to whether the links are direct or occur due to other effects to which there is a common connection. In the case of cosmic rays, the correlation with TSI is undoubtedly due to effects of magnetic fields, which effect both cosmic rays and the TSI. Surface magnetic fields cause variations in the TSI, due to the creation of sunspots and faculae, and a small fraction of these magnetic fields can propagate into the outer heliosphere and thus modulate cosmic rays entering the solar system. The cosmic ray connection would almost certainly also explain the link between TSI and the coronal source flux. It is also important to better understand how the variation of sunspot spread is linked to the

emergence rate, as this is a directly measurable quantity from the Greenwich observations. If a correct link can be established it should be possible to directly calculate the emergence rate from the Greenwich observations.

Another area of that needs serious investigation is the possibility that heat flux blocked by the effects of magnetic fields in the convection zone, can return to the solar surface after a period of time which is determined by the duration of the magnetic flux blockage. This returned heat flux is dependent upon the duration and amount of magnetic flux causing heat flux blockage, thus as sunspot activity increases so will the amount of heat flux being blocked. This heat flux could then return at a later date and thus could contribute to the increase in solar minimum irradiance that is believed to be causing the long-term increases in TSI that have been found. Thus this possible effect needs to be investigated further as it could prove extremely important in irradiance reconstructions.

Alternatively the long-term variations could be due to small, unresolved flux in magnetogram data. However, it should be possible to place limits on this "hidden flux" effect. In a magnetogram pixel, in and out flux would partially or totally cancel giving $|B/\mu|$ near zero, but there would be brightening associated with both the inward and outward field. The survey of *Ortiz et al.* [2003] does not show much scatter in contrast, which would result from this effect. However the *Ortiz et al.* study does not extend below a threshold of $|B/\mu| = 17G$. Investigation of the contrast at smaller $|B/\mu|$ is needed if we are to quantify the effect of brightening caused by the flux that is hidden in magnetograms.

The reconstructions I created of TSI calculated from the coronal source flux need further investigation in the light of my findings of the minimum possible values of TSI during the Maunder minimum. The same applies to other reconstructions of TSI based on stellar analogues. It is important that a correct and robust reconstruction of TSI is developed so that long-term variations in the global climatic system can be accurately measured. To this end it is important that the effect of faculae on the TSI can be modelled effectively, because the data series of PSI is reliable and this thesis has shown that we can extrapolate PSI back to 1874, or even 1702, with some confidence.

The 3 models of TSI developed form the equations of *Ortiz* in chapter 7, need to be investigated as to their accuracy and possible impacts upon the terrestrial climate. In particular, the contrasts I have used derived by *Ortiz et al.* [2002] are for a wavelength of λ=676.8 nm, the continuum emission observed by MDI which is in the red part of the visible spectrum. *Unruh et al.* [1999] have studied the spectrum of facular contrasts and although this red line gives a contrast close to the average value (relevant to TSI), in general a correction factor is needed which might be a factor of μ.

The Sun impacts upon the terrestrial climate in a number of ways, for example either directly through its irradiance or indirectly via the modulation of cosmic rays. The understanding of these effects is still in its infancy, and it is clear that climate system needs to be understood in far greater detail. The Sun's influence will not work in isolation and other influences on climate (anthropogenic, volcanoes) will combine with solar variations in a complex and non-linear manner. Plans are advanced for the TSI reconstructions derived and presented here to be input into the Hadley centre's global coupled models (HAD3CM and it's successors) to evaluate their implications for climate change. The Hadley centre is part of the UK Meteorological Office, responsible for studies of global climate change. This is an example of the greater cooperation between the disciplines of environmental science and solar-terrestrial physics, which is needed if the role of solar variability on climate is to be properly untangled from the other influences.

Index of terms:

ACRIM: Active Cavity Radiometer for Irradiance Monitoring

A_f: Facular area

A_G: Sunspot group area.

Carrington rotation: Carrington rotation 1 began on November 9, 1853 and a new rotation begins each time the prime meridian of the Sun (i.e. where the heliographic longitude is zero) passes the centre point on the solar disk as seen from earth.

ERBS: Earth radiation Budget Satellite.

F_b: Facular brightening.

F_{ba}: Active region facular brightening.

F_{bn}: Network facular brightening.

F_{bno}: Facular enhancement to the quiet sun.

HF: Hickey-Frieden Radiometer, onboard the Nimbus 7 satellite.

MDI: Michelson Doppler Imager, onboard SoHO.

PMOD: Physikalisch-Meteorologisches Observatorium Davos.

PSI: Photometric Sunspot Index.

Q_o: magnetic free quiet sun irradiance.

Q_s: Quiet sun irradiance at solar minimum.

SoHO: Solar and Heliospheric observatory.

TSI: Total Solar Irradiance.

VIRGO: Variability of Solar Irradiance and Gravity Oscillations.

References:

Allen, A., Astrophysical Quantities, 3rd. Edition, Athlone Press, London, 1973.

Bard, E., G. Raisbeck, F. Yiou, and J. Jouzel, Solar irradiance during the last 1200 years based on cosmogenic nuclides, *Tellus, 52B*, 985-992, 2000.

Baranyi, T., L. Gyoeri, A. Ludmany and H.E. Coffey, Comparison of sunspot area data bases, Mon. Not. R. Astron. Soc., 323, 223-230, 2001.

Beer, J., A. Blinov, G. Bonani, H.J. Hofman, R.C. Fintel, Use of BE-10 in polar ice to trace the 11-year cycle of solar activity, *Nature*, 347, 1990.

Beer, J., S. Tobias, and N. Weiss, An active Sun throughout the Maunder minimum, *Sol. Phys.*, 181, 237-249, 1998.

Berger, T.E., M.G. Lofdahl, G. Scharmer, A.M. Title, Observations of magnetic convection in sunspots with 100km resolution, 34th meeting of the solar physics devision of the American Astronomical society, 2003.

Bering, E.A., A.A. Few, and J.R. Benbrook, the global electric circuit, *Phys. Today,* 51, 24-30, 1998.

Bhatnagar, A., K. Jain, and S.C. Tripathy, Variation of solar irradiance and mode frequencies during Maunder minimum, *Astro-ph*/0201025, 2002.

Bierman, L., *Vierteljahrsschr. Ast. Ges., 76,* 194, 1941.

Bond, G., B. Kromer, J. Beer, R. Muscheler, M.N. Evans, W. Showers, S. Hoffman, R. Lotti-Bond, I. Hajdas, G Bonani, Persistent solar influence on North Atlantic climate during the Holocene, *Science, 294*, 2130-2136, 2001.

Chapman, G. A., *J. Geophys. Res., 92*, 809, 1987.

Chapman, G. A., Private communication 2003.

Chapman, G. A., A. M. Cookson and J. J. Dobias, Solar variability and the relation of facular to sunspot areas during solar cycle 22, *The Astrphys. Journal*, 482, 541-545, 1997.

Chapman, G. A., A. M. Cookson, J. J. Dobias, and S. R. Walton, An Improved Determination Of The Area Ratio Of Faculae To Sunspots, *The Astrophysical Journal*, 555:462-465, 2001.

Cliver, E.W., and A.G. Ling, Secular change in geomagnetic indices and the open solar flux in the first half on the twentieth century, *J. Geophys. Res., 107(A10), paper* SSH11, 2002.

Cliverd, M. A., Clark T. D. G., Clarke E., and Rishbeth H., Increased magnetic storm activity from 1868 to 1995, *J. Atmos. Sol-Terr. Phys.*, 60, 1998.

Deinzer, W., G. Hansler, M. Schussler, E. Weisshar, *Ap J.* 139, 435, 1984.

Fligge, M. and S. K. Solanki, Inter-cycle variations of solar irradiance: sunspot areas as a pointer, *Sol. Phys.*, 173, 427-439, 1997.

Fligge, M. and S. K. Solanki, Long-term behaviour of emissions from solar faculae: Steps towards a robust index, *Astron. Astrophs.*, 332, 1082-1086, 1998.

Fligge, M., S.K. Solanki, Y.C. Unruh, C. Fröhlich and Ch. Wehrli, A model of solar total and spectral irradiance variations, *Astron. Astrophys.*, *335*, 709-718, 1998.

Foster, S. and M. Lockwood, Long-term changes in the solar photosphere associated with changes in the coronal source flux, *Geophys. Res. Lett.*, 28, 8, 1443-1446, 2001.

Foukal P., The behaviour of solar magnetic plages measured from Mt. Wilson observations between 1915-1984. *Geophys. Res. Lett.*, 23, 16, 1996.

Foukal, P., A comparison of variable solar total and ultraviolet irradiances in the 20th century, *Geophys. Res. Letts.*,29, 2002.

Foukal, P. and L Milano, A measurement of the quiet network contribution to solar irradiance variation, *Geophys. Res. Lett., 28*, 883-886, 2001.

Foukal, P., K. Harvey, and F. Hill, Do changes in the photospheric magnetic network cause the 11-year variation of total solar irradiance?, *Astrophys. J., 383*, L89-L92, 1991.

Fröhlich C., observations of irradiance variability, Space Sci. Rev. 94, 2000.

Fröhlich C., Long-term behaviour of space radiometers, Metrologia, 40, 2003.

Fröhlich, C., Solar irradiance variations, Proc. ISCS-2003, ESA-SP-535, 183-193, ESA publications, ESTEC, 2003a

Fröhlich, C. and W. Finsterle, VIRGO radiometery and total solar irradiance in 1996-2000 revised, in *"Recent insights into the physics of the Sun and the heliosphere: highlights from SOHO and other space missions", IAU Sumposium, 203,* 105-110, 2001.

Fröhlich, C. and J. Lean, The Sun's total irradiance: cycles, trends and related climate change uncertainties since 1976, *Geophys. Res. Lett.,* 25, 4377-4380, 1998.

Fröhlich, C. and J. Lean, Total solar irradiance variations, in new eyes to see inside the sun and stars, ed. F. L. Deubner et al. Proc. IAU symp., 185, Kyoto, August 1997, 1998a.

Fröhlich, C., J. M. Pap, and H. S. Hudson, Improvement of the photometric sunspot index and changes of the disk-integrated sunspot constant with time, *Sol. Phys.*, 152, 111-118, 1994.

Gray, D.F., and W.C. Livingstone, Monitoring the solar temperature: spectroscopic temperature variations of the sun, *Astrophys. J., 474*, 802-809, 1997.

Hale G. E., F. Ellerman, S. B. Nicholson, A. H. Joy, ApJ, 49, 153, 1919.

Harvey, K.L., in "The Sun as a variable star: solar and stellar irradiance variations", eds. J.M. Pap, C. Fröhlich, H.S. Hudson, S.K. Solanki, *IAU Col, 143,* 217-225, Cambridge Univ. Press, 1994.

Harvey, K.L., The Solar Activity Cycle and Sun-As-A-Star Variability in the Visible and Infrared, in Solar Analogs: Characteristics and Optimum Candidates, Proc. 2nd annual Lowell Observatory Fall Workshop, October 5-7, 1997, ed. Ed. J.C. Hall, 1997 (http://www.lowell.edu/users/jch/workshop/sa.html).

Harvey, K.L. and H.S.Hudson, Adv Space Res., 25, 1735, 2000

Harvey, K.L., and C. Zwaan, Properties and emergence of bipolar active regions, *Sol. Phys.*, 148, 85-118, 1993.

Hathaway, H., Private communication, 2004.

Hoyt, D., and K. Schatten, A discussion of plausible solar irradiance variations 1700-1992, *J. Geophys. Res.,* 98, 18, 895-18, 906, 1993.

Hudson, H.S., S. Silva, M. Woodward, and R.C. Willson, The effects of sunspots on solar irradiance, *Solar Physics, 76*, 211-219, 1982.

Kivrova, N.A., S.K. Solanki, and M. Fligge, Total solar magnetic flux: dependence on spatial resolution of magnetometers, i*n "From Solar Min to Max: Half a solar cycle with SoHO"*, Proc. SoHO 11 Symposium, Davos, Switzerland, March

2002, ESA-SP-508, pp 155-158, ESA Publications, Noordvijk, The Netherlands, 2002.

Kivrova, N.A., S.K. Solanki, M. Fligge, and Y.C. Unruh, Reconstruction of solar irradiance variations in cycle 23: is solar surface magnetism the cause?, *Astron. Astrophys.*, 383, 706-712, 2003.

Knaack, R., M. Fligge, SW.K. Solanki and Y.C. Unruh, The influence of an inclined rotation axis on solar irradiance variations, *Astron. and Astrophys., 376*, 1080-1089, 2001.

Knölker, M., M. Schussler, E. Weisshaur, Model calculations of magnetic flux tubes. III-Properties of solar magnetic elements, *Astron. & Astrophys.*, 194, 1988.

Krimigis, S.M., R.B. Decker, M.E. Hill, T.P. Armstrong, G. Gloeckler, D.C. Hamilton, L.J. Lanzerotti, E.C. Roelof, Voyager 1 exited the solar wind at a distance of ~85 AU from the Sun, *Nature*, 426, 45 – 48, 2003.

Kuhn, J.R. and K.G. Libbrecht, Non-facular solar luminosity variations, *Astrophys. J.,* 381, L35-L37, 1991.

Kuhn, J.R. and R.F. Stein, Accounting for the solar acoustic and luminosity variations from the deep convection zone, *Astrophys. J.,* 463, L117-L119, 1996.

Kuhn, J.R., R.I. Bush, X. Scheick, and P. Scherrer, The Sun's shape and brightness, *Nature 392,* 155-157, 1988.

Kuhn, J.R., H. Lin, and R Coulter, What can irradiance measurements tell us about the solar magnetic cycle?, *Adv. in Space Res., 24(2)*, 185-194, 1999.

Lean, J., Evolution of the Sun's spectral irradiance since the Maunder minimum, *Geophys. Res. Lett.,* 27 , 2425-2428, 2000a.

Lean, J., Short-term direct indices of solar variability, *Space Sci. Rev.*, 94, 39-51, 2000b.

Lean, J., J. Beer and R. Bradley, Reconstruction of solar irradiance since 1610, implications for climate changes, *Geophys. Res. Lett.*, 22, 3195-3198, 1995.

Lean, J.L., Y.-M. Wang and N. R. Sheeley Jr., The effect of increasing solar activity on the Sun's total open magnetic flux during multiple cycles: Implications for solar forcing of climate, *Geophys. Res. Lett.*, 29, 24, 2002.

Lockwood, M., Long-term variations in the Magnetic Fields of the Sun and the Heliosphere: their origin, effects and implications, *J. Geophys. Res.*, 27, 2425-2428, 2001.

Lockwood, M., An evaluation of the correlation between open solar flux and total solar irradiance, *Astron. Astrophys.*, 382, 678-687, 2002a.

Lockwood, M., The Relationship between the near-Earth interplanetary field and the coronal source flux: Dependence on timescale, *J. Geophys, Res.*, 107, 2002b.

Lockwood, M., Long-term variations in the open solar flux and links to variations in earth's climate, *in "From Solar Min to Max: Half a solar cycle with SoHO"*, Proc. SoHO 11 Symposium, Davos, Switzerland, March 2002, ESA-SP-508, pp 507- 522, ESA Publications, Noordvijk, The Netherlands, 2002c.

Lockwood M., Twenty-three cycles of changing open solar flux. J.of Geophys. Res. 108, 2003.

Lockwood, M., R.B. Forsyth, A. Balogh and D. J. McComas, The accuracy of open solar flux estimates from near-Earth measurements of the interplanetary magnetic field: analysis of the first two perihelion passes of the Ulysses spacecraft, *Annales Geophys.*, in press, 2004.

Lockwood, M. and S. Foster, Long-term variations in the magnetic field of the Sun and possible implications for terrestrial climate, *in "The solar cycle and Terrestrial Climate"*, Proc. 1st. Solar and Space Weather Euroconference, ESP SP-463, 85-94, 2001.

Lockwood, M., and R. Stamper, Long-term drift of the coronal source magnetic flux and the total solar irradiance, *Geophys Res. Lett., 26*, 2461-2464, 1999.

Lockwood, M., R. Stamper and M. N. Wild, A doubling of the sun's coronal magnetic flux during the last 100 years, *Nature*, 399, 437-439, 1999a.

Lockwood, M., R. Stamper, M.N. Wild, A. Balogh and G. Jones, Our changing sun, *Astron & Geophys*, *40*, 4.10-4.16, 1999b.

Mackay D.H., E.R. Priest, M. Lockwood, The Evolution of the Sun's open magnetic flux: I. A single bipole. Solar Phys. 2002.

Mackay, D. H and M. Lockwood. The Evolution of the Sun's Open magnetic Flux: II. Full solar cycle simulations, *Solar Physics*, *209(2),* 287-309, 2002.

Marsh, N., and H. Svensmark, Cosmic rays, clouds and climate, Space Sci. Rev. 2000.

Mayaud, P.N., The aa indices: A 100-year series characterising the magnetic activity, J. Geophys. Res., 77, 6870-6874, 1972.

McDonald, F.B., E.C. Stone, A.C. Cummings, B. Heikkila, N. Lal, and W.R. Webber, Enhancements of energetic particles near the heliospheric termination shock *Nature*, 426, 48 – 51, 2003.

Mendoza, B., *Astrophys. J., 483*, 523, 1997.

Neckel, H., and D. Labs, Solar limb darkening 1986-1990 (λλ303 to 1099nm), Solar Physics, 153: 91-114, 1994.

Nesme-Ribes, E., E.N. Ferreira, R. Sadournay, H. Le Truet and Z.X. Li, Solar dynamics and its impact on solar irradiance and terrestrial climate, *J. Geophys. Res., 98*, 18923-18935, 1993.

Nindos, A., and H. Zirin, The relation of Ca II K features to magnetic field, Solar Physics, 179, 253-286, 1998.

Ortiz A., V. Domingo, B. Sanahuja and L. Sánchez, An Example of Isolated Active Region Energy Evolution: NOAA 7978 in Proc. 1st Solar and Space Weather Euroconference: "The Solar Cycle & Terrestrial Climate", ESA SP-463, p.395, 2000.

Ortiz, A., V. Domingo, B. Sanahuja and C. Fröhlich, Excess facular emission from an isolated active region during solar minimum: the example of NOAA AR 7978, *J. of Atmospheric and Solar-Terrestrial Physics*, accepted (2003)

Ortiz A., S.K. Solanki, V. Domingo, M. Fligge, and B. Sanahuja, On the intensity contrast of solar photospheric faculae and network elements, *Astron. & Astrophys.* 388, 1036-1047, 2002.

Ortiz, A., S.K. Solanki, M. Fligge, V. Domingo, and B. Sanahuja, On the contrast of faculae and small magnetic features, *in "The solar cycle and Terrestrial Climate", Proc. 1st. Solar and Space Weather Euroconference, ESP SP-463*, 399, 2001.

Parker, E.N., The dynamical oscillation and propulsion of magnetic fields in the convective zone of a star. I. General considerations, *Astrophys. J., 312*, 415-420, 1995.

Parker, E.N., Theoretical properties of Ω-loops in the convective zone of the Sun. II. The origin of enhanced solar irradiance, Astrophys. J., 440, 415-420, 1995.

Pulkkinen, P.J., J. Brooke, J. Pelt, and I. Tuominen, Long-term variation of sunspot latitudes, *Astron. Astrophys., 341*, L43-L46, 1999.

Pulkkinen, T.I., H. Hevanlinna, P.J. Pulkkinen and M. Lockwood, The Earth-Sun connection in time scales from years to decades to centuries, Space Sci. Rev., 2000.

Pulkkinen, T.I., H. Nevanlinna, P.J. Pulkkinen, and M. Lockwood, The Earth-Sun connection in time scales from years to decades to centuries, *Space Sci. Rev.,* 95, 625-637, 2001.

Rast, M. P., P. A. Fox, H. Lin, B. W. Lites, R. W. Meisner, O. R. White, Bright rings around sunspots, *Nature, 401*, 678 – 679, 1999.

Reid, G.C., Solar forcing and the global climate change since the mid-17th century, *Climatic Change, 37*, 391-405, 1997.

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

Russell, C.T., R.L. McPherron, Semi annual variation of geomagnetic activity, *J. Geophys. Res.*, 78, 1973.

Sanchez-Cuberez, M., M. Vazquez, J.A. Bond and M. Sobotka, Centre to limb variation of solar granulation in the infrared, *Astron and Astrophys*. 397, 2003.

Schatten, K.H., H.G. Mayr, K. Omidvar, and E. Maier, A hillock and cloud model for faculae, *Astrophys.. J., 311*, 460-473, 1986.

Scherrer P.H., R.S. Bogart, R.I.Bush, J.T. Hoeksema, A.G. Kosovichev, J. Schou, W. Resenberg, L. Springer, T.D. Tarbell, A. Title, C.J. Wolfson, I. Zayer and the MDI Engineering Team, The Solar Oscillations Investigation – The Michelson Doppler Imager, *Solar Physics, 162*, 129-188, 1995.

Schrijver C.J. C. Zwann, Solar and stellar magnetic activity, Cambridge university press, 2000.

Schröder, P., R. Smith and K. Apps, Solar evolution and the distant future of the Earth, *Astron. & Geophys.*, 42, 2001.

Solanki S.K., and V. Brigljević, Continuum brightness of solar magnetic elements, *Astron. Astrophys., 262*, L29-L32, 1992.

Solanki, S.K. and M. Fligge, Solar irradiance since 1874 revisited, *Geophys. Res. Lett., 25*, 341-344, 1998.

Solanki, S.K. and M. Fligge, A reconstruction of total solar irradiance since 1700, *Geophys. Res. Lett.,* 26,1999.

Solanki, S. K. and M. Fligge, Reconstruction of past solar irradiance, *Space Science Reviews, 94*, 127-138, 2000.

Solanki, S. K. and M. Fligge, How much of the solar irradiance variations is caused by the magnetic field at the solar surface?, *Adv. In Space Res.,* 29(12), 1933-1940, 2002.

Solanki, S.K., M. Schüssler, and M. Fligge, Secular evolution of the Sun's magnetic field since the Maunder minimum, *Nature, 480*, 445-446, 2000.

Solanki, S.K., M. Schüssler, and M. Fligge, Secular variation in the Sun's magnetic field, *Astron. Astrophys.*, 383, 706-712, 2002.

Solanki, S.K. and J.O. Stenflo, Properties of solar magnetic flux tubes as revealed by Fe1 lines, *Astron. Astrophys., 140*, 185-198, 1984.

Spruit, H. C., Pressure equilibrium and energy balance of small photospheric fluxtubes, *Solar Phys.*, 50, 1976.

Spruit, H. C., Appearance at the solar surface of disturbances in the heat flow associated with differential rotation, *Astron. and Astrophys.*, 55, 1977.

Spruit, H. C., The flow of heat near a sunspot. *Astron. and Astrophys.*, 108, 1982a

Spruit, H. C., Effect of spots on a stars radius and luminosity, *Astron. and Astrophys.*, 108, 1982b.

Spruit, H.C., Theory of luminosity and radius variations, *in "The Sun in Time"*, Eds. C. P. Sonnet, M. S. Giampapa and M. S. Matthews, Univ. of Arizona Press, pp. 118-159, ISBN 0-8165-1297-3, 1991.

Spruit, H.C., Theory of solar irradiance variations, *Space Sci. Rev., 94*, (1/2), 113-126, 2000.

Stamper, R., M. Lockwood, M.N. Wild, and T.D.G. Clark, Solar Causes of the Long-Term Increase in Geomagnetic Activity, *J. Geophys Res., 104,* 28325-28342, 1999.

Steiner O., U. Grossmann-Doerth, M. Schussler, M. Knolker, proceedings of an international Workshop on solar polarization, *Solar Phys*, 164, 1996.

Stott, P. A., et al., External control of 20th century temperature by natural and anthropogenic forcings, *Science,* 290, 2133-2137, 2000.

Svalgaard, L., E. W. Cliver and P. Le Sager., No doubling pf the sun's coronal magnetic field during the last 100 years. Submitted 2003.

Svensmark, H., and E. Friis-Christesen, Variation of cosmic ray flux and global could coverage: A missing link in solar-climate relationships, J. Atmos. Sol. Terr. Phys., 59, 1997.

Svensmark, H., Cosmic rays and Earth's climate, *Space Science reviews*, 93, 2000.

Svensmark, H., Influence of cosmic rays on Earth's climate, Phys. Rev, Lett, 81, 1998.

Taylor R. J., The Sun as a Star. Cambridge University Press, 1997.

Tett, S. F. B, et al., Causes of twentieth century temperature change near the Earth's surface. *Nature*, 399, 569-572, 1999.

Tolstoy, L. What men live by, and other stories, The coffee-house of Surat.

Topka, K. P., T. D. Tarbell, and A. M. Title, Properties of the smallest solar magnetic elements II. Observations versus hot walls models of faculae, *Astrophys. Journal.* 484, 1997.

Unruh, Y.C., S. K. Solanki, and M. Fligge, The spectral dependence of facular contrast and solar irradiance variations, Astron. Astrophys., 3456, 635-642, 1999.

Usoskin, I. G., S. Solanki, M. Schüssler, K. Mursula and K. Alano, Millenium-scale sunspot number reconstruction: Evidence for an unusually active sun since the 1940s, *Phys. Rev. Lett.*, 91, 21, 2003.

Walton, S.R., D.G. Preminger, and G.A. Chapman, The Contribution of Faculae and Network to Long Term Changes in the Total Solar Irradiance, *Astrophys. J*, in press, 2003.

Wang, Y.-M., N.R. Sheeley, Jr., and J. Lean, Understanding the evolution of Sun's magnetic flux, *Geophys. Res. Lett., 27*, 621-624, 2000a.

Wang, Y.-M., J. Lean, and N.R. Sheeley, Jr., The long-term evolution of Sun's open magnetic flux, *Geophys. Res. Lett., 27*, 505-508, 2000b.

Wang, Y.-M., J. Lean, and N.R. Sheeley, Jr., Role of a variable meridional flow in the secular evolution of the sun's polar fields and open flux, *Astrophys. J.*, 2002.

Wang. T.-M, and N.R.Sheeley, Jr., Sunspot activity and the long-term variation of the Sun's open magnetic flux, *J. Geophys. Res.*, 2002.

Wenzler, T., S.K. Solanki, D.M. Fluri, C. Frutiger, M. Fligge and A. Ortiz, Modelling solar irradiance variations: separate models for the network and active region faculae, i*n "From Solar Min to Max: Half a solar cycle with SoHO"*, Proc. SoHO 11 Symposium, Davos, Switzerland, March 2002, ESA-SP-508, pp 231-234, ESA Publications, Noordvijk, The Netherlands, 2002.

Wenzler, T., Private communication, 2004.

Wigley. T. M. L., and S. C. B. Raper, Climatic changes due to solar irradiance changes, *Geophys. Res. Lett.*, 17, 12, 2169-2172, 1990.

Willson R., Total solar irradiance trend during solar cycle 21 and 22, Science, 277, 1997.

Wilson PR, Altrock RC, Harvey KL, Martin SF, Snodgrass HB, The extended solar activity cycle, *Nature*, 333, 748, 1988.

Willson, R.C., H.S. Hudson, and G.A. Chapman, Observations of solar irradiance variability, *Science, 211*, 700-702, 1981.

Willson R., and Mordvinov A., Secular Total solar irradiance trend during solar cycles 21-23, 2002.

Wolff, C.L. and J.R. Hickey, Solar irradiance change and special longitudes due to r-modes, Science, 235, 1631-1633, 1987a

Wolff, C.L. and J.R. Hickey, Multiperiodic irradiance change caused by r-modes and g-modes, *Sol. Phys., 109*, 1-18, 1987b

Zhang, Q., W.H. Soon, S.L. Baliunas, G.W. Lockwood, B.A. Skiff, and R.R. Raddick, A method of determining possible brightness variations of the Sun in past centuries from observations of solar-type stars, *Astrophys. J., 427*, L111-L114, 1994.