Mem. S.A.It. Vol. 76, 781

Memorie della

Mechanisms of secular magnetic field variations

S. K. Solanki and M. Schüssler

Max-Planck-Institut für Sonnensystemforschung, Max-Planck-Str. 2, 37191, Katlenburg-Lindau, Germany; e-mail: solanki-office@mps.mpg.de

Abstract. The variability of the solar magnetic field on time scales of decades and longer lies at the root of the various mechanisms by which the changing Sun could affect Earth's climate. We discuss the origin of the secular variability of both the open heliospheric flux and the total unsigned solar surface flux and review models that have been put forward to describe these variations. We propose that a combination of the effects of overlapping activity cycles and the long decay time of large-scale magnetic patterns is responsible for the secular variability of the solar magnetic field.

Key words. Sun: activity – Sun: faculae, plages – Sun: irradiance – Sun: magnetic fields – solar-terrestrial relations – sunspots – Sun: UV radiation

1. Introduction

The Sun's variability is increasingly invoked as a driver of the Earth's climate. Now, solar variability manifests itself in many ways, not all of them of equal relevance for the climate. The most discussed variable parameters of the Sun of possible importance for the Earth's climate system are:

1. The total solar irradiance, which affects the energy input into the climate system (e.g., Fröhlich 2000; Cubasch & Voss 2000).
2. The solar spectral irradiance, in particular in the UV, which affects stratospheric chemistry that could in turn influence tropospheric properties (e.g., Haigh, this issue; Rozanov et al., this issue).
3. The Sun's open magnetic flux, which modulates the galactic cosmic ray flux. The galactic cosmic rays have been proposed to influence cloud cover and hence climate (e.g., Svensmark & Friis-Christensen 1997).

All of these quantities (and many more) are influenced by the structure and evolution of the Sun's magnetic field. In the first two cases, the total magnetic flux of the Sun (in the second, mainly the flux outside sunspots) is the driver of the variability; in the third case, only the open magnetic flux is important.

The magnetic field, created by a dynamo process in the solar interior, most likely located near the interface between the convection zone and the radiative core, (see Schmitt, this issue, and Charbonneau 2005) is transported to the solar surface by buoyancy and appears there in the form of bipolar magnetic regions. The largest of these are well visible as large active regions, which harbour sunspots surrounded by faculae. These active regions are only the tip of the iceberg, however. Smaller bipolar regions are more abundant, with the number of such regions appearing on the surface increasing with area to the power -2 (Schrijver & Harvey 1994). The smaller active regions only have small dark pores and are mainly composed of faculae (i.e. smaller magnetic features). Still smaller are the ephemeral regions, which are more broadly distributed over the Sun and individually carry a much smaller magnetic flux (typically a few times 10^{19} Mx) than the larger active regions. In spite of the lower amount of magnetic flux per feature, the total amount of magnetic flux transported to the solar surface by ephemeral regions, $2–4 \cdot 10^{26}$ Mx/yr (Hagenaar 2001) is orders of magnitude larger than the flux transported by the larger active regions, $3 \cdot 10^{23}–3 \cdot 10^{24}$ Mx/yr (Harvey 1993).

Very soon after its emergence at the surface, the magnetic flux in an activce region becomes organized into sunspots, pores and magnetic elements that to a certain extent evolve as discrete entities. At the same time, the flux starts to evolve and to decay again. The evolution is driven by the interaction of the field with convection (mainly granulation and supergranulation in layers near the solar surface), solar differential rotation and large scale flows (mainly meridional circulation), the decay is driven by the resistivity of the plasma, with reconnection and ohmic dissipation playing an important role. The magnetic flux is buffeted by the convective flows and performs a random walk over the solar surface. On a larger scale, this can be described as a diffusion process. Opposite-polarity fields are brought together, so that they can reconnect and dissipate. As a consequence, active regions decay, with a part of the magnetic field in the region decaying away in situ through processes associated with cancellation with opposite polarity field. However, part of the magnetic flux diffuses out into the surrounding quiet Sun, where it forms the network (initially an enhanced network, later the normal quiet network). In addition to this source, the network is also fed by magnetic flux emerging from the solar interior in the form of ephemeral regions.

Most of the Sun's flux is closed at a given time, while a small fraction is open in the sense that the field lines reach out into the heliosphere, including interplanetary space. These open fields largely come from regions with a dominant magnetic polarity (unipolar regions) and are often (but not exclusively) associated with coronal holes. There are also open field lines emanating from active regions.

The number of bipolar regions appearing at the solar surface changes cyclically along with the number of sunspots. At sunspot minimum, very few large active regions appear at the solar surface and there are times when not a single active region is present, while at activity maximum many active regions are to be seen at any given time. For the smaller bipolar regions, the contrast between activity maximum and minimum is not so strong. For the ephemeral regions, Harvey (1993) found that their appearance rate changes by a factor of only 2 between activity maximum and minimum (compared to over an order of magnitude for large active regions and sunspots). She also found that ephemeral regions appeared over an extended cycle, which starts before the sunspot cycle (Harvey 2000). For the smallest observed ephemeral regions there is almost no cycle dependence of the emergence rate and even an indication of a slightly anticyclic behavior (Hagenaar et al. 2003).

What is the evidence that the Sun's magnetic field exhibits a secular change? Solar activity does vary on time scales longer than a cycle. For example, no two solar cycles are identical. The record of sunspots reveals strong and weak cycles (with evidence for a modulation of cycle strength by a roughly 90 year period, the Gleissberg period) and intervals with practically no visible cyclic activity at all (e.g., during the Maunder minimum in the second half of the 17th century, Eddy 1976). Such a change in cycle properties need not necessarily cause a secular change in the magnetic field, however. Sunspot numbers, for example, do not exhibit a background level that would vary slowly with time.

The first indications for a possible secular change in the Sun's magnetic properties came from observations of cool stars. Baliunas & Jastrow (1990) studied the distribution of the radiative flux in the cores of the Ca II H and K spectral lines in field stars. This quantity is related via a power law to the magnetic flux in faculae (e.g., Schrijver et al. 1989). Baliunas & Jastrow found that the Ca II distribution shows two peaks, a broad peak at stronger calcium emission levels and a narrower peak at very low emission levels. The Sun is found to reside only in the upper half of the upper peak. They interpreted this distribution as distinguishing between stars with and without cyclic activity (inhabiting the upper and lower peak, respectively).

White et al. (1992) and Lean et al. (1992) used this result in order to estimate the difference in solar irradiance between the Maunder minimum and today. Actually, they found that the change in irradiance deduced from the stellar observations was so large that it could not be completely accounted for by changes of the observed magnetic field.

Unfortunately, recent observations (Wright 2004; Giampapa 2005) of larger and in particular more carefully selected samples of stars reveal a single-peaked distribution, with cycling and non-cycling stars being intermixed, i.e. showing no significant offset in Ca II H & K flux relative to each other. One explanation for the discrepancy between these more recent results and those of Baliunas & Jastrow (1990) may be that the latter authors combined stars of different ages in their plot. In particular, many of the non-cycling stars may actually have already evolved off the main sequence and thus cannot be directly compared with the Sun in a Maunder minimum-like state.

Just as doubts were beginning to appear regarding the value of the stellar data as a guide to secular variations of the solar magnetic field, a new line of evidence emerged. Lockwood et al. (1999) made a reconstruction of the interplanetary magnetic field (i.e. the part of the open solar magnetic flux lying near the ecliptic) based on the geomagnetic *aa* index, which has been recorded since 1868. They found that the interplanetary magnetic field, besides showing a cyclic variation, also exhibited a strong secular trend and in particular showed roughly a doubling in the course of the last century. During the time that direct, space-based measurements are available (since 1964) they agree relatively well with the reconstructions. Measurements made by the Ulysses space probe have also shown that the strength of the radial component of the Sun's open magnetic field is roughly independent of the heliographic latitude, so that the reconstructed interplanetary field also provides a good measure of the open flux of the Sun (e.g., Lockwood 2002).

Note that although only a small part of the Sun's total flux is in the form of open field lines, these data are particularly important since the record of total (unsigned) surface flux reaches back over a far shorter period of time. Now, the amount of open flux can change due to the rearrangement of magnetic polarities on the solar surface without any change in the total amount of magnetic flux, a secular change in total or spectral irradiance, however, requires a change in the amount of total magnetic flux at the solar surface (assuming that the surface magnetic field is responsible for all solar irradiance changes). Since at solar activity minimum the field is dominantly contained in the magnetic network, a change in the total magnetic flux implies a secular change in the strength of the magnetic network. Consequently, there would be fewer or weaker network elements associated with, e.g., weaker cycles or during grand minima, such as the Maunder minimum.

2. Explanation of secular change

So far, the only mechanisms proposed to produce a secular change of the Sun's magnetic flux are driven by overlaps between consecutive solar activity cycles, in the sense that the flux appearing in one cycle is still present when flux in the next cycle starts to appear.

The overlap can be produced in two ways:

1. Magnetic flux survives at the solar surface for a sufficiently long time, so that some flux from the previous activity cycle still remains when sunspots (i.e., active regions) belonging to the next cycle start appearing in large numbers.
2. The appearance of magnetic bipoles of the following cycle starts while bipoles belonging to the old cycle are still appearing. Bipoles can be assigned to a given cycle by Hale's polarity law, although this becomes increasingly difficult for small bipoles. Alternatively, they can be assigned

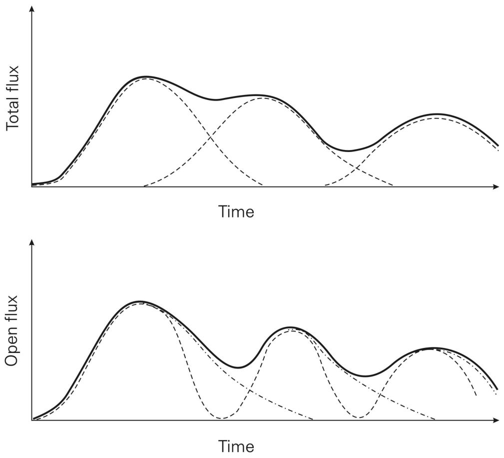

Fig. 1. Illustration of basic mechanisms for producing a background magnetic field exhibiting secular change. Upper panel: overlapping cycles. Shown is a sketch of the time series of total magnetic surface flux over parts of three activity cycles. The total amount of flux emerging in each cycle (dashed lines) is assumed to decay instantaneously (i.e., on time scales short compared with the length of the cycle). The solid curve gives the sum of the fluxes due to each cycle. Lower panel: long decay time of open flux. The dashed line gives the emerging flux in large active regions (scaled to the same level as the open flux). The dot-dashed lines show the open flux due to each cycle (scaled to the same level as the active-region flux), including a long decay time. The resulting total amount of open flux (sum of the dot-dashed curves) is given by the solid line.

by the latitudes at which they most commonly emerge.

The first mechanism was proposed by Solanki et al. (2000) to explain the secular change in the Sun's open magnetic flux, but it probably plays a role also for other large-scale patterns of magnetic field. The second mechanism was suggested by Solanki et al. (2002) as a means of providing a secular change in the Sun's total magnetic flux. The extended emergence time of ephemeral active regions leads to such an overlap between cycles).

Fig. 1 illustrates the basic principles. As long as flux that appeared (or is appearing) in the previous cycle is still present when flux in the new cycle starts to emerge, the amount of (open as well as total) flux on the Sun does not drop to zero at activity minimum, but remains at some finite level. This mechanism explains in a natural manner the rather large amount of open magnetic flux present at solar activity minimum and the significant strength of the network at all phases of the solar cycle (i.e. the significant amount of total surface flux always present).

3. Results of a coarse model

Using a simple model, Solanki et al. (2000, 2002) could reproduce the reconstruction of the open flux by Lockwood et al. (1999), including the secular increase of flux over the last century (with a single free parameter, the decay time of the open flux). This is illustrated in Fig. 2. By employing the observations of Harvey (2000) of an overlap between cycles due to the early emergence of ephemeral regions, Solanki et al. (2002) could also predict a secular increase in the total magnetic flux at the solar surface. In this case, however, the uncertainties are larger since there is insufficient information on the behaviour of ephemeral regions, which have a strong effect on the total flux but only weakly affect the open flux. In both cases, the magnetic flux has a background component that shows a secular trend.

The variation of the total magnetic flux calculated from the model can be compared with observations carried out at Mount Wilson Observatory (MWO), Wilcox Solar Obervatory (WSO), and National Solar Observatory (NSO). We must, however, keep in mind that the relatively low spatial resolution of the synoptic charts on which the plotted data are based miss a significant amount of flux in the quiet Sun due to the mixing of polarities on relatively small scales. Krivova & Solanki (2004) estimated that at least half the true flux is not detected due to this effect. After taking this effect into account (assuming that flux in the quiet Sun is a factor of 3 larger than measured), we can compare the observed and modelled total magnetic flux. This comparison is shown in Fig. 3. The computed flux fits the observed values well within the scatter of the individual data points. Note that the large difference between cycles

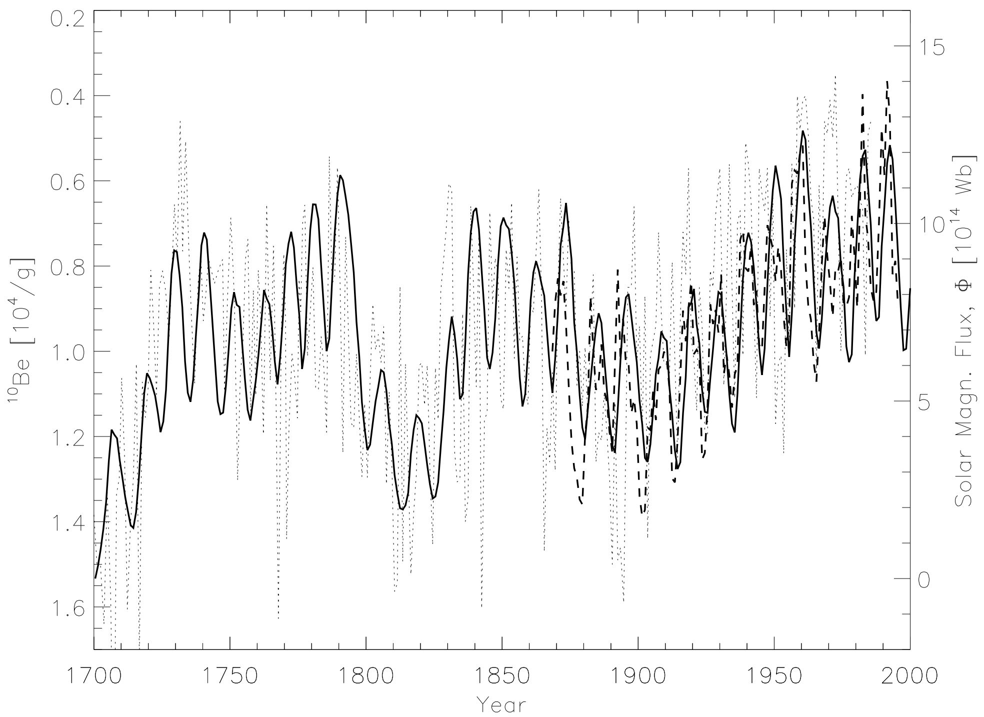

Fig. 2. Time series since 1700 of open solar magnetic flux computed using a coarse model (solid curve), the interplanetary magnetic field reconstructed by Lockwood et al. (1999, dashed curve) and ^{10}Be concentration in the Dye-3 ice core (Greenland; Beer et al. 1990, dotted curve). Adapted from Solanki et al. (2000).

21 and 22, which led to a mismatch between the model predictions and the NSO data set in an earlier comparison (Solanki et al. 2002) is not visible here. This is partly due to the fact that the other data sets do not show the same difference between the two cycles and partly because the NSO data during large parts of cycle 21 have been shown to be too low (Arge et al. 2002; Wenzler et al. 2005).

The irradiance record resulting from this magnetic field reconstruction is discussed by Krivova & Solanki (this issue).

4. Flux transport models of magnetic field evolution

Flux transport models solve the equation describing the transport of magnetic flux at the solar surface, i.e. the two-dimensional induction equation on a spherical surface. This provides a more realistic description of the magnetic field evolution than the coarse model discussed in the previous section.

In a flux transport model, the times and positions of emergence, the initial sizes, and the orientations of magnetic bipoles describing large active regions are specified. These data can be taken (statistically or directly) from observations. For instance, the times, numbers and latitudes of emergence can be taken from the butterfly diagram of sunspots, limiting the emerging flux to that in active regions with sunspots. The size distribution can be taken from the observations of Schrijver & Harvey (1994), who showed that the number N of active regions with a certain area A follows $N(A) \sim A^{-2}$. The magnetic field resulting from the emerged bipoles is then allowed

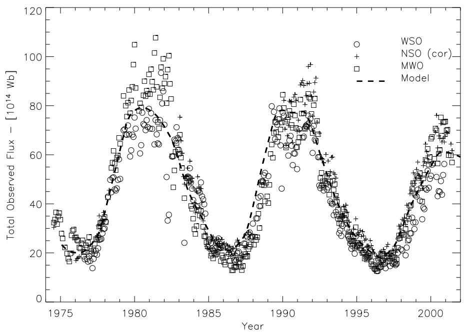

Fig. 3. Comparison between measured and computed total solar magnetic flux. Circles: observations made at Wilcox Solar Observatory (WSO); squares: observations made at Mount Wilson Observatory (MWO); crosses: observations made at the National Solar Observatory, after corrections introduced by Arge et al. (2002). The dashed line is the result of the coarse model (from Balmaceda et al., in preparation).

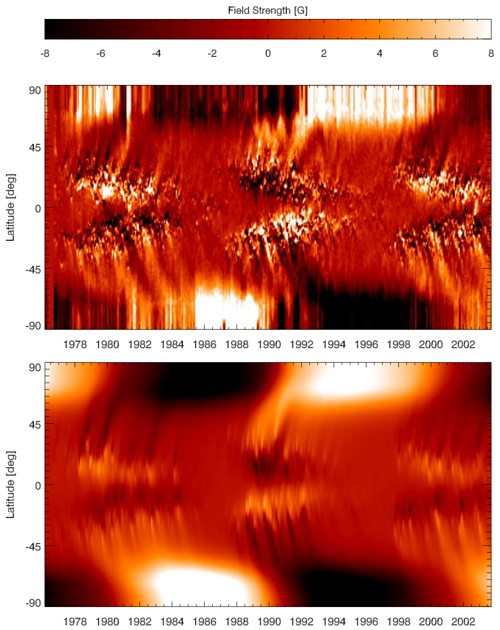

Fig. 4. Comparison between observations and flux transport model computations. Upper panel: Time-latitude diagram of the azimuthally averaged radial photospheric magnetic field through the period 1976–2005 created from NSO synoptic charts (courtesy David Hathaway, NASA/MSFC). Dark and bright shading corresponds to opposite magnetic polarities. Lower panel: Time-latitude diagram resulting from the flux transport model of Baumann et al. (2004).

to evolve according to the two-dimensional induction equation.

The flux emerging in ephemeral regions, although in principle also of relevance, is generally not explicitely included in the flux transport models because the emphasis of these models is on the large-scale structures to which the ephemeral regions do not contribute significantly, owing to their small sizes and their large scatter in emergence latitudes and tilt angles.

Sinks of magnetic flux also need to be prescribed. The physics of flux removal from the solar surface is complex and not fully understood. Flux transport codes greatly simplify this part of the physics. Sinks of magnetic flux are simply described by flux cancellation provided by the diffusion term.

Drivers of the flux evolution are convection, which is described by the (turbulent) diffusion term in the induction equation, differential rotation, and meridional circulation. While the latter two are known from observations, the value of the turbulent diffusivity is a free parameter. A value of the order of 600 $\mathrm{km^2{\cdot}s^{-1}}$ is determined by starting such a code from the magnetographic observations at a given time and comparing the results after a few solar rotations with the updated observations.

Flux transport models reproduce the observed time-latitude diagrams of the azimuthally averaged magnetic field rather well if the input distribution of bipolar regions is similar to the butterfly diagram of sunspots (e.g., Wang et al. 1989; van Ballegooijen et al. 1998; Mackay et al. 2002; Baumann et al. 2004). Fig. 4 shows the observed time-latidude diagram for the period 1974–2005 (upper panel) in comparison with the result from the flux transport model of Baumann et al. (2004, lower panel).

Flux transport computations confirm the basic mechanism proposed to explain secular

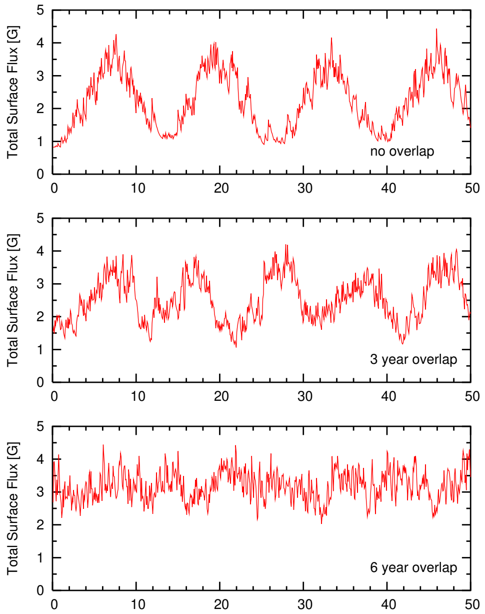

Fig. 5. Total magnetic flux vs. time produced by test runs with a flux transport model. Top panel: No overlap between the emergence of flux in consecutive cycles; middle panel: 3 years of overlap; bottom panel: 6 years of overlap.

changes in the open and total flux, namely that the level of the total magnetic flux at activity minimum (background magnetic flux) depends on the overlap between consecutive cycles and on the strength of these cycles. Baumann et al. (2004) considered a set of identical cycles, with different amounts of overlap between consecutive cycles. Fig. 5 gives the total surface flux for an overlap of 0 years, 3 years and 6 year (from top to bottom). Clearly, while the peak level of the total flux (i.e. the unsigned flux integrated over the whole Sun) remains approximately the same in all cases, the level of minimum total flux values rises dramatically with increasing overlap time, reaching almost the same value as the maximum for an overlap of 6 years, so that the signature of the underlying cycles is practically wiped out.

One particularly interesting feature visible in Fig. 5 is that even for no overlap between the flux emerging in successive cycles (top panel), a significant background flux is built up and consistently maintained. This goes beyond what is predicted by the coarse model discussed in Sect. 3, which only produces such a background when an overlap is present. The reason is simple. In the coarse model, all active region fields have a single decay time (generally taken to be 4 months), while in the flux transport models different spatial scales of the magnetic flux pattern decay at different rates. Small scales decay faster, larger scales slower. This is a consequence of the second spatial derivative in the diffusion operator. Turbulent diffusion and systematic flows tend to spread the magnetic flux over large areas and the remaining large-scale patterns decay more slowly. This leads to an overlap between the field of successive regions even if there is no overlap in the emergence. Such a background field is also subject to secular change (Baumann et al. 2004).

Using the sunspot number to mark emerging flux (i.e., assuming no overlap between the emerging flux of successive cycles) Schrijver et al. (2002) and Wang et al. (2005) have found that the open flux computed according to their recipes agrees relatively well with the reconstruction of Lockwood et al. (1999). It also agrees relatively well with the simple model of Solanki et al. (2000). Note that the level of the open flux is only slightly affected by the introduction of smaller active regions not harbouring sunspots (such as ephemeral regions), since these contribute only little to the open flux (but have a significant effect on the total flux).

5. Summary

A mechanism is available to explain and compute the secular change of the Sun's open and total magnetic flux (and the interplanetary magnetic field). The underlying concept is that of overlapping solar cycles, i.e. magnetic field that emerged in one cycle is still present when fresh flux starts emerging in the new cycle. The overlap can be produced by one or both of two processes:

1. A long decay time (a year or more) of part of the magnetic flux.
2. Emergence of magnetic flux belonging to the new cycle (e.g. in the form of

ephemeral regions) before the end of the previous cycle.

The basic results of a coarse model are confirmed by more detailed flux transport computations. At least for the case when the emergence strictly follows the times when sunspots are present, the flux transport models give an even higher level of the background flux.

The computed open flux agrees well with the best current observations Lockwood et al. (1999) and with indirect measures of open flux, such as the concentration of cosmogenic isotopes like ^{10}Be (Usoskin et al. 2002).

The observational constraints on the total magnetic flux are much weaker, since regular, reliable, direct observations are available for less than 40 years. Therefore there is a strong need for observations that constrain the level of secular change of the Sun's total magnetic flux.

Acknowledgements. We thank L. Balmaceda, I. Baumann, and N. Krivova for providing material and for helpful discussions.

References

Arge, C. N., Hildner, E., Pizzo, V. J., & Harvey, J. W. 2002, J. Geophys. Res. (Space Physics), 107, DOI 10.1029/2001JA000503
Baliunas, S. & Jastrow, R. 1990, Nature, 348, 520
Baumann, I., Schmitt, D., Schüssler, M., & Solanki, S. K. 2004, A&A, 426, 1075
Beer, J., Blinov, A., Bonani, G., Hofmann, H. J., & Finkel, R. C. 1990, Nature, 347, 164
Charbonneau, P. 2005, Living Reviews in Solar Physics, LRSP-2005-2, http://solarphysics.livingreviews.org
Cubasch, U. & Voss, R. 2000, Space Sci. Rev., 94, 185
Eddy, J. A. 1976, Science, 192, 1189
Fröhlich, C. 2000, Space Sci. Rev., 94, 15
Giampapa, M. 2005, in The Sun, Solar Analogs and the Climate, ed. I. Rüedi, M. Güdel, & W. Schmutz (Berlin: Springer), 307–415
Hagenaar, H. J. 2001, ApJ, 555, 448
Hagenaar, H. J., Schrijver, C. J., & Title, A. M. 2003, ApJ, 584, 1107
Harvey, K. 2000, in Encyclopedia of Astronomy and Astrophysics, ed. P. Murdin (Bristol & London: Institute of Physics & Nature Publishing Group)
Harvey, K. L. 1993, PhD thesis, University of Utrecht, The Netherlands
Krivova, N. A. & Solanki, S. K. 2004, A&A, 417, 1125
Lean, J., Skumanich, A., & White, O. 1992, Geophys. Res. Lett., 19, 1595
Lockwood, M. 2002, in ESA SP-508: From Solar Min to Max: Half a Solar Cycle with SOHO, 507–522
Lockwood, M., Stamper, R., & Wild, M. N. 1999, Nature, 399, 437
Mackay, D. H., Priest, E. R., & Lockwood, M. 2002, Sol. Phys., 209, 287
Schrijver, C. J., Cote, J., Zwaan, C., & Saar, S. H. 1989, ApJ, 337, 964
Schrijver, C. J., DeRosa, M. L., & Title, A. M. 2002, ApJ, 577, 1006
Schrijver, C. J. & Harvey, K. L. 1994, Sol. Phys., 150, 1
Solanki, S. K., Schüssler, M., & Fligge, M. 2000, Nature, 408, 445
—. 2002, A&A, 383, 706
Svensmark, H. & Friis-Christensen, E. 1997, Journal of Atmospheric and Terrestrial Physics, 59, 1225
Usoskin, I. G., Mursula, K., Solanki, S. K., Schüssler, M., & Kovaltsov, G. A. 2002, jgr, 107, 13
van Ballegooijen, A. A., Cartledge, N. P., & Priest, E. R. 1998, ApJ, 501, 866
Wang, Y.-M., Lean, J. L., & Sheeley, N. R. 2005, ApJ, 625, 522
Wang, Y.-M., Nash, A. G., & Sheeley, N. R. 1989, ApJ, 347, 529
Wenzler, T., Solanki, S. K., & Krivova, N. A. 2005, A&A, 432, 1057
White, O. R., Skumanich, A., Lean, J., Livingston, W. C., & Keil, S. L. 1992, Publ. Astron. Soc. Pac., 104, 1139
Wright, J. T. 2004, AJ, 128, 1273