1 Introduction

The occurrence of sunspots has long been observed to wax and wane with an approximately 11-year periodicity. The same cyclic behaviour is present in nearly all observed manifestations of solar magnetic activity, from small-scale ephemeral regions to coronal mass ejections, while solar cycle effects have been measured from the deep solar interior to the outer edges of the heliosphere. Through modulation of solar wind structures and energetic particle populations, the solar cycle directly influences space weather. It is also the most accessible diagnostic of the solar (and hence a stellar) magnetic dynamo. The history and properties of the solar cycle are first discussed in terms of sunspots, before summarising solar cycle variations in properties observed remotely and in situ. Longer-term variations, on the centennial and millennial scales, are also presented.

2 Sunspot number

The solar cycle is synonymous with periodic variations in sunspot occurrence. While sunspots are only an indirect proxy for the solar magnetic field central to the solar cycle, they are by far the longest series directly observed solar parameter and thus remain of great scientific interest across a range of research areas.

Telescopic observations of sunspots began in the early 17th century, but early observers (with a few notable exceptions) tended to make relatively intermittent measurements and over a relatively limited period of time. The scattered nature of the data, as well as limitations in analysis and visualization, meant that the approximately 11-year cycle in sunspot number – so obvious to us now – was not definitively identified for nearly 250 years. Christian Horrebow, a Danish astronomer, first suggested in 1776 that the number of sunspots may vary periodically, but that the available observations are not sufficient to identify such a signature [36]. Heinrich Schwabe, a German astronomer, undertook 18 years of dedicated solar observations, summarised in Figure 1. He reported an approximately 10-year period between the maxima in annual numbers of
Figure 1: Heinrich Schwabe’s sunspot and auroral observations from 1826-1844. Top: The annual number of sunspot groups (black) and sunspot-free days (red). Data have not been corrected for the number of observing days per year, though the trends are unchanged. Bottom: Observed aurora per year.
sunspot groups and sunspot-free days [74, 3]. Hence the sunspot cycle is often referred to as the Schwabe cycle.

Schwabe’s result was not widely known in the scientific community of the time, but Rudolf Wolf, director of the Bern Observatory, recognised its importance and established a formal programme of daily sunspot observation. These data were compiled into a “relative” sunspot number, $R$, which is not simply the total number of sunspots on the solar disc, but the number of individual spots plus ten times the number of sunspot groups [94, 95]. This definition means $R$ is dominated by the number of groups, which are generally easier to observe than individual spots with smaller telescopes and earlier optics, though determining what does/doesn’t constitute a group adds some level of subjectivity. In order to enable data from multiple observers (or the same observers with different telescopes or techniques) to be combined, Wolf further introduced a scaling factor ($k$). In this way, Wolf effectively began the process of collating and combining all available sunspot observations which continues today. Known observational logs allow $R$ to be reconstructed back to around 1750 at monthly resolution and to 1700 at annual resolution (data are available from http://www.sidc.be/silso/datafiles). The record of solar activity can be extended further back in time, to 1610, by producing annual means of only sunspot groups [41], though as for $R$ some means of calibration between observers is required.

Recently, the interpretation of historical sunspot records, the methodologies used to combine different observers and counting techniques have come under renewed scrutiny [14, 80]. On-going coordinated efforts [13, 15] are bringing together new statistical techniques, greater physical understanding of the processes which determine sunspot occurrence and newly discovered historical records, to correct for systemic observer biases and produce more robust composite sunspot series [12, 88, 82, 50, 11].

The various methods to produce composite sunspot series (Figure 2) display a number of common features. Most striking is the Maunder minimum (MM), a period of greatly reduced solar activity, first identified by a reduction in auroral occurrence [21, 85], but also clearly present in sunspot occurrence and other solar proxies. The MM spans approximately 1650 to 1715, though this depends somewhat on the chosen threshold and sunspot record. There is evidence from solar activity proxies [64], that the solar cycle continued at some level despite the absence of a clearly identifiable Schwabe cycle [89]. A smaller and shorter “grand minimum” of activity, the Dalton minimum (DM), spans the two Schwabe cycles of the early 19th century. All sunspot records also show a minimum in Schwabe cycle magnitudes around the start of the 20th century, followed by a rise until the largest cycle in the 1950s, after which cycle magnitudes have declined. These features highlight that the Sun undergoes long-term variability, superimposed on the approximately 11-year solar cycle. Such variations are discussed further in Section 5.

While the solar cycle is traditionally thought of as an 11-year cycle, this is only true on average. Figure 3 shows the cumulative distribution function of Schwabe cycle lengths over the interval 1700–present. The median of the
Figure 2: Annual sunspot number over the last four centuries. Top: Relative (Wolf) sunspot number, $R$, from SILSO (Sunspot Index and Long-term Solar Observations World Data Center, Royal Observatory of Belgium) [12]. The latest version (V2, red, released July 2015) incorporates a number of corrections over the previous version (V1, black), which has been scaled by a factor 1.43 to produce agreement in the most recent cycles and enable easy comparison. Bottom: The number of sunspot groups from three composite series. Black shows the original [41] series, red shows the [82] series and blue shows reconstruction from [88]. In both panels, grey-shaded areas show alternate solar cycles back to 1700. Two periods of reduced solar activity are highlighted; the Dalton Minimum (DM) and the Maunder Minimum (MM).
Figure 3: The cumulative distribution function of Schwabe cycle lengths, in years. The median cycle length (solid black lines) is 10.5 years, while the interquartile range (black dashed lines) spans 10 to 12 years. The minimum and maximum observed cycle lengths are 9 and 14 years, respectively.

distribution is 10.5 years, while the interquartile range spans 10–12 years. The minimum and maximum of the distribution are 9 and 14 years, respectively. Note, however, that there remains some debate about whether the 14-year cycle (cycle 4, spanning 1784 to 1799) is actually two smaller cycles [87]. Excluding this value reduces the maximum to 13 years.

Figure 4 shows the sunspot number variation as a function of solar cycle phase [63]. This allows the variation between cycles to be compared, accounting for the differing cycle lengths. The top panels show that there is a great deal of variability (approximately a factor 3) in the magnitude of the cycles, but that the waveform is reasonably repeatable. This is further evidenced by the bottom panels, where each cycle has been scaled to the cycle mean. It is clear that the rise and decline of cycles is asymmetric, with the smoothed peak occurring approximately 35–40% of the way through the cycle [90].

A number of relations between solar cycle features have been reported [36]. The best known is that the sunspot number rise time and the peak sunspot number are anti-correlated (i.e., larger cycles show a faster rise) [90]. More recent analysis [17], however, suggests this “Waldmeier effect” may not be statistically robust.

Perhaps just as important as the sunspot number is the spatial distribution of spots. The top panel of Figure 5 shows that the time variation of sunspots in the northern and southern hemispheres is often slightly decoupled. The phase shift in hemispheric activity may possibly give rise to the “Gnevyshev gap” [28], the drop in total sunspot number (and a number of other disc-averaged properties,
Figure 4: Top-left: SILSO V2 monthly sunspot number as a function of solar cycle phase for 24 solar cycles. Top-right: Mean SSN across all cycles (red) and the standard deviation within cycles (pink). The bottom panels show same data scaled by the mean SSN of each cycle.
Figure 5: Sunspot variations as recorded by the Royal Greenwich Observatory (RGO) from 1874 to 1976, after which it was continued by the US Air Force’s Solar Optical Observing Network (SOON). Daily data are used to construct 3 Carrington rotation (81 day) averages. Top: The number of sunspot groups visible on the whole Sun (black) and in the northern (red) and southern (blue) hemispheres. Middle: Number of sunspot groups as a function of latitude and time, called the butterfly diagram and also referred to as Spörer’s law. The saturated colourbar shows the average number of groups per day at a given latitude. Bottom: The average latitude of sunspot groups.
discussed further in Section 3) around the peak of the solar cycle. (However, it has also been suggested that double peak features exist within single hemisphere variations [60].) The gap feature is particularly prominent in the most recent solar cycle (approximately 2012-2014), when the initial peak in total sunspot number was primarily a result of northern-hemisphere activity, while the later surge in activity was primarily a southern-hemisphere phenomenon.

The mean latitude and latitudinal spread of sunspots vary systematically over the solar cycle, which was first noted by Carrington [10] and further analysed by Spörer [79]. The middle panel of Figure 5 shows that the start of a solar cycle sees sunspots confined to two latitudinal bands centred around 20-25° either side of the solar equator. As the solar cycle progresses, these bands slowly drift to lower latitudes, reaching approximately 5-11° of the equator by the end of the cycle [53]. As can be seen in the bottom panel of Figure 5, this behaviour is largely invariant of solar cycle magnitude or length [77] and the sharp increase in the average latitude of sunspots is a useful means of defining the start of a new solar cycle [63].

There is typically some degree of temporal overlap in emergence of high-latitude sunspots of a new solar cycle with the tail end of the low-latitude sunspots of the old solar cycle. This leads to the overlapping “butterfly diagram” shown in Figure 5 which may be linked to an “extended solar cycle” seen in other magnetic proxies at higher latitudes [93, 56].

Of course, the value of sunspot observations lies in their use as a visible proxy for the photospheric magnetic field, which was first directly measured by George Ellery Hale using the Zeeman effect [33]. Sunspots were found to occur in bipolar pairs. In each solar hemisphere, the leading sunspot of each pair (with respect to solar rotation) has the same magnetic polarity over the whole solar cycle, an observation now referred to as Hale’s law [34]. This polarity is opposite in each hemisphere and is the same as the polar magnetic field polarity at the start of the solar cycle (see Figure 6). Thus, after the polar field reversal around sunspot maximum, the polarity of the leading sunspot opposes the polar field in each hemisphere (see Figure 6). The magnetic orientation of spot pairs reverses each solar cycle, meaning a complete magnetic cycle of the Sun comprises two Schwabe cycles (i.e., approximately 22 years).

Also shown in Figure 6 is Joy’s law: The observation that, on average, the leading sunspot (with respect to the rotation direction) in each pair is at lower latitude than the trailing sunspot [35]. The tilt angle of the sunspot pairs is ≈ 7° and, on average, increases with latitude [92].

The three main laws of solar cyclic activity, Hale’s and Joy’s laws as well as Spörer’s butterfly diagram, provide fundamental observational constraints for theoretical models of solar cycle, i.e. dynamo models from the first conceptual model [6] to full 3-dimensional numerical models [9]. Their combined effect, coupled with convective (granular and supergranular) flows and the bulk flows like differential rotation and meridional circulation in the photosphere and convective zone, are the building blocks of the solar dynamo, which drives the solar cycle [6, 9].
Figure 6: A schematic of Hale’s and Joy’s laws of sunspot orientations combined with the equatorward shift of sunspot locations expressed in Spörer’s butterfly diagram. Red and blue shaded areas indicate inward (negative) and outward oriented (positive) photospheric magnetic fields, respectively. Sunspots occur in pairs, with the leading spot at lower latitude (Joy’s law). In each hemisphere, the magnetic polarity of the leading spot is the same as the dominant polarity at the start of the solar cycle (Hale’s law). Note that in the declining phase of the cycle the sunspot pairs are closer to the equator than during the rise phase.
3 Beyond sunspots: The magnetic solar cycle

3.1 Magnetograms

The most important discovery in solar physics in the early 20th century was the discovery of magnetic fields in sunspots by Hale [33], which –besides spectroscopy– brought physics into solar astronomy and transformed it to solar physics. Hale could only measure strong magnetic fields, however. It was only in the 1950s when with the magnetograph built by Harold and Horace Babcock [7] that weak (of a few gauss field strength) magnetic fields outside sunspots could be measured, providing evidence of a global magnetic field, and most importantly, of time-evolving polar fields [5], which were foreseen by Hale. Horace Babcock discovered that polar fields are evolving with time and even reverse in polarity.

The weak magnetic fields outside sunspots were found to be ubiquitous in the solar photosphere and became just as important as sunspot fields in understanding the solar cycle. There is a continuity from the formation of strong flux concentrations through the emergence of magnetic flux bundles from the solar interior (which create bipolar pairs of sunspots), their decay through flux dispersion driven by convective motions and magnetic cancellation processes, the migration of flux towards the poles driven by a poleward meridional flow, which modulates and reverses the polar fields.

Long-term, continuous observations of the photospheric magnetic field have revealed a strong solar cycle variation in the total unsigned photospheric flux [2]. However, this translates to a much smaller variation in total unsigned flux at the top of the corona, the open solar flux (OSF) [91]. See also Section 4.

Examples of magnetic maps covering the solar surface (combining a central strip from daily magnetograms during a full solar rotation) are shown in Figures 7 and 8, in Carrington Rotations 2111 (June–July 2011) and 2194 (August–September 2017), prior to and after the polarity reversal of the solar poles. Note the strong bipolar magnetic fields (tilted to the solar equator as described by Joy’s law) and that the polarity orientation of bipolar pairs is different in the north and south hemispheres (Hale’s law), cf. Figure 6. Furthermore, note the dispersed weaker magnetic fields around the stronger polarity concentrations. The dispersed fields have long poleward extensions (polarity streams) bent by differential rotation (i.e. that the poles rotate slower than the lower latitudes and the equator). Also notice that in 2011 strong magnetic fields were at higher latitudes than in 2017, which is consistent with Spörer’s butterfly diagram.

Stacking longitudinal averages of full-surface synoptic magnetic maps (a full map summed up in one pixel column) builds up the so-called magnetic butterfly diagram (Figure 9), which is a key to tracking the build-up, decay and polarity change of polar fields. Each half of the butterfly wings shows the polarity of the leading spots in the corresponding hemisphere (Hale’s law) due to an asymmetry in field strength between the leading and following spots in the sense of the solar rotation, as well as Joy’s law, (i.e. that the leading and following polarities are not at the same solar latitudes). It is also clear from the magnetic butterfly diagram that most of the poleward-moving streams originating from each butterfly
Figure 7: Magnetic field map of the entire solar surface in Carrington Rotation 2111, June–July 2011, around the first peak of solar cycle 24 and prior to the polarity reversal of the solar poles. Red- and blue-shaded areas (with orange and green in between, see colour bar) indicate inward (negative) and outward oriented (positive) photospheric magnetic fields (G), respectively. Note the strong bipolar magnetic fields (tilted to the solar equator as described by Joy’s law) and that the polarity orientation is different on the north and south hemispheres (Hale’s law), cf. Figure 6. Furthermore, note the dispersed weaker magnetic fields around the stronger polarity concentrations. The dispersed fields have long poleward extensions (polarity streams) bent by differential rotation, i.e. that the poles rotate slower than the lower latitudes.
Figure 8: Magnetic field map of the entire solar surface in Carrington Rotation 2194, August–September 2017, when the waning cycle 24 had its likely final local peak and after the polarity reversal of the solar poles. Red- and blue-shaded areas (with orange and green in between, see colour bar) indicate inward (negative) and outward oriented (positive) photospheric magnetic fields (G), respectively. Note that in the declining phase of the cycle the sunspot pairs are closer to the equator than during the maximum shown in Figure 7.
wing are of opposite polarity to that of the wing itself, i.e. the dominant polarity of the polewards streams is that of the following polarity of bipolar sunspot groups. This fact has its origins in the tilt of the bipolar pairs, Joy’s law. The following polarity spots are closer to the poles, therefore their fields, while getting dispersed by granular and supergranular motions, have a higher probability to make it to the poles. Through magnetic cancellation, the poleward-moving streams weaken the polar field strength, eventually reversing the polarity of the poles. After the reversal, which occurs around solar maximum, the poleward-moving streams are of the same polarity as the polar field and start building up new polar flux which reach their maximum at the end of each cycle. At the same time, the lower-latitude leading polarities have a higher probability to make it to the solar equator, where they meet and cancel with opposite-polarity leading fields from the opposite hemisphere. Joy’s law contributes another way to the disparity between the migration pattern of leading and following polarities: The tilted division line between the two polarities, the so-called magnetic polarity inversion line, means that poleward diffusing leading polarities get cancelled by the following/trailing polarity and vice versa. The poleward polarity streams do not always carry the “correct” magnetic flux, however. Active regions that disobey Joy’s law may produce occasional opposite-polarity streams and disturb the polar flux cancellation or build-up process. Such a large “rogue” active region resulted in the negative-polarity stream on the northern hemisphere in early 2011 in cycle 24 [97]. Scatter in tilt angles of bipolar sunspot pairs leads to a variability of the polar fields, which play a key role in the solar cycle, as they serve as seed fields in dynamo models and therefore modulate cycle amplitudes and can even trigger Grand (Maunder-minimum-like) Minima [58].

In Figure 9 it is also clear that the polarity change of the poles takes place around solar maximum. This may not happen at the same time on the two hemispheres, and there can be several months and even more than a year difference between them. The polar fields are the strongest just before the start of solar minimum. Note the weak polar fields from about 2007 onward.

3.2 Solar Dynamo

Horace Babcock in 1961 made a synthesis of the complex observational data known at the time, incorporating them into an ingenious model of the 22-year solar cycle [6]. For details and modern dynamo models we refer the reader to another chapter in the Oxford Research Encyclopedia of Physics dedicated to the solar dynamo [9]. Here, we briefly summarise the key elements of a flux transport dynamo model which is mainly based on Babcock’s conceptual model and Robert Leighton’s kinematical model that includes magnetic flux transport [47], but also includes the effects of a poleward meridional circulation [20, 84] and results from helioseismology on the latitudinal and radial differential rotation [73, 38] which were not known to Babcock and Leighton. The most recent models, which numerically solve the so-called dynamo equation [68] and incorporate modern advances in magneto-hydrodynamics, share the most fundamental elements with the early models.
Figure 9: Magnetic butterfly diagram of the last four solar cycles 20–24 (courtesy D. Hathaway). Each (pixel) column is a full-Sun synoptic map (like the ones shown in Figures 7 and 8) the sum of magnetic fields along a given (pixel) latitude. Note that each half of the butterfly wings shows the polarity of the leading spots in the corresponding hemisphere (Hale’s law). Note the opposite-polarity poleward streams originating from each butterfly wing, which weaken the polar fields and eventually, around the time of solar maximum, change the magnetic polarity of the poles. The polar fields are the strongest at the end of each cycle when the strongest magnetic fields (spots) are close to the solar equator. Note the weak polar fields from about 2007 onward.
Figure 10: Cartoons showing the progression of the cycle based on the Babcock–Leighton dynamo mechanism [72].

At solar minimum, when the polar fields are the strongest, the global magnetic fields are considered to be purely poloidal: The Sun has a big bar-magnet-like bipolar field (Figure 10, leftmost cartoon). In the interior of the Sun, plasma $\beta$, a ratio of plasma and magnetic pressures, is high (i.e. plasma motions deform/carry magnetic fields and not the other way around like in the solar corona). Therefore differential rotation, which extends into the Sun nearly radially down to the bottom of the $\approx 200,000$ km-deep convection zone, will carry magnetic fields faster around the equator than at higher latitudes. This increasingly shears the magnetic field and gradually transforms the poloidal fields into a helical spiral around the equator, i.e. forms a toroidal field. This is called the $\Omega$-effect (Figure 10, first three cartoons). Babcock placed the toroidal field (i.e. the dynamo layer) at a shallow depth of 0.1 solar radii, but it is now rather thought to be located at the bottom of the convection zone.

Winding up the toroidal field leads to its amplification. Babcock also involved twisting by a roller-bearing effect by radial differential rotation to further amplify the sub-photospheric field (such radial differential rotation shear is indeed present at the bottom of the convection zone). The latitudinal structure of amplification is determined by the differential rotation’s latitudinal profile. The amplification proceeds fastest at high latitudes of the activity belt: around $\pm 30^\circ$. Deep in the Sun in the toroidal-field “dynamo layer”, magnetic fields (flux tubes) have lower plasma density than their surrounding field-free plasma, which at some point makes them buoyant [67]. This will lift some flux tube strands out of the general toroidal layer and make them rise to and eventually through the surface of the Sun. Flux emerging from the general toroidal field will satisfy Hale’s law on both hemispheres (bipolar pairs will be oppositely oriented). As amplification proceeds to lower latitudes, or the so-called dynamo wave progresses equatorward, which is thought to be assisted by the deep-seated
reverse flow of the meridional circulation, flux emergence will gradually shift to lower latitudes, which is consistent with Spörer’s butterfly diagram. Flux tubes or magnetised plasma parcels rising through the convective zone of decreasing gas pressure expand and as they rise in a rotating medium, they are subjected to the Coriolis force, which makes them deflect towards the poles and rotate clockwise in the northern and anti-clockwise in the southern hemisphere, leading to a tilt of emerging sunspot pairs consistent with Joy’s law. Helical turbulence in convective motions makes rising magnetic elements carry a poloidal field component appropriate to the next cycle (Figure 10, α-effect).

The decay and dispersal of the emergent magnetic flux leads to the neutralization and subsequent reversal of the general poloidal field as a result of the systematic inclination/tilt of sunspot pairs (i.e. that the following polarity tends to lie at higher latitude than the leading polarity: Joy’s law). The dispersing following-polarity fields have a higher probability to reach the poles, while leading polarity fields are more likely to reach the equator. When opposite polarities are brought together due to the dispersal of magnetic fields from different sunspot regions, equal amounts of opposite flux cancel. As a result, about 99% of emergent flux cancels against remnant flux from neighbour sunspot pairs, less than 1% of following polarity makes it to the nearest pole, first neutralizing the existing fields and then replacing them by flux of opposite polarity. The same fraction of leading polarities of the two hemispheres cancel in the equatorial strip. The poleward surface meridional flow (of ≈ 20 m/s) assists the poleward migration of magnetic concentrations. The flow carries magnetic flux of both polarities, however, due to Joy’s law and the consequent orientation of the magnetic inversion line between leading and following polarities on both hemispheres, poleward-moving leading magnetic polarities have a higher probability to be cancelled as they diffuse into their dispersing opposite-polarity counterparts. We therefore see mainly following-polarity poleward streams in the magnetic butterfly diagram (Figure 9).

The poleward-dispersing following polarities weaken and eventually neutralise and reverse the polar fields. The meridional circulation which carried the field poleward also will carry it down to the base of the convection zone via turbulent diffusion. Changes in the meridional circulation are thought to affect the strength of the polar fields.

By the end of the cycle the Sun has oppositely oriented polar fields to those it started with (Figure 10, rightmost cartoon); it will take the next cycle to get back to the original poloidal field orientation. This is why the magnetic or Hale solar cycle length is 22 years.

As the polar fields serve as seed fields for the succeeding solar cycle, the strength of polar fields are linked to the amplitude of the succeeding cycle. The weak polar fields we see Figure 9 in the period 2007–2009 during the solar minimum were followed by a weak cycle 24. The polar field strengths presently building up at the solar poles seem to indicate another weak cycle 25.
3.3 Helioseismology

Helioseismology, which provides us with means of observing plasma motions in the invisible interior of the Sun, has shown that differential rotation extends into the convective zone, and that in the radiative zone below, the Sun’s rotation becomes rigid-body like. These two zones are separated by a narrow zone of radial shear, called the tachocline, where the large-scale dynamo is thought to operate. Helioseismology has also revealed that the differential rotation profile is not constant, but is modulated by migrating bands of a few meters per second faster and slower rotation, a phenomenon known as torsional oscillation [37], which extends about 0.1 solar radii into the solar interior [1]. At lower latitudes, the $\approx 10^\circ$ wide bands move toward the equator with time. The greatest concentration of sunspot pairs is associated with the poleward edge of the equatorward-moving band. At higher latitudes, however, the bands move poleward, which is similar to the observed poleward movement of magnetic streams at high latitudes [1], and indeed these features match quite well [39]. Acceleration in the surface rotation at a given latitude usually can be seen in advance of the appearance of magnetic activity (emergence of sunspot pairs) [44], and this even can forecast the appearance of a new cycle. During cycle 24, the weak poleward branch indicated a slower overall rotation at high latitudes [40], which may be related to the observed weaker polar fields [70].

3.4 Electromagnetic spectrum

As solar magnetic fields heat the solar atmosphere, their cyclic variability influences the Sun’s electromagnetic output across the spectrum. The integrated output, or total solar irradiance (TSI), was originally referred to as the “solar constant.” However, sensitive space-based measurements of TSI confirmed a small (< 0.1%) solar cycle variation [25, 32], as shown in Figure 11. There has been much debate about the level of the TSI during the Maunder minimum, but recent reconstructions have converged on a consensus of around only 0.1 to 0.2 % lower than the space age [45]. The small amplitude of the TSI variation means solar variability has only a minor effect on global terrestrial warming in comparison with other known forcings [30, 78].

TSI is obviously dominated by emission in the visible and infrared wavelengths associated with the photosphere, where dark and cool sunspots decrease the TSI, while bright and hot facular regions, which form above dispersing sunspot magnetic fields, increase the TSI [98]. From these competing effects the influence of faculae is twice as strong, resulting in an increase of spectral irradiance in most wavelength ranges with increasing solar activity. However, the relative contribution to the TSI by sunspots and faculae changes during the lifetime of an active region: during its formation, the effects of cool sunspots dominate, while during its decay phase the influence of hot facular regions becomes dominant. As the emergence phase of active regions is shorter than their decay phase, at solar maximum, when there are many active regions of various ages on the Sun, the influence of hot faculae is stronger. Proxies developed to
model the contribution of strong and weak magnetic fields to the TSI [42] are the Mount Wilson Sunspot and Magnetic Plage Strength Indices, the fractional area coverage of the solar surface by magnetic fields stronger and weaker than 100 G (but > 10 G). The CaII K line emission can also be used as a proxy for magnetic flux, and therefore shows solar cycle variability [22]. The MgII index, derived from the ratio of the MgII h and k lines to the neighbouring continuum, can also be used as a proxy for facular contribution to the TSI [46].

Shorter wavelengths result from emission higher in the solar atmosphere. Extreme ultraviolet (EUV) emission, which is dominated by the coronal extension of strong magnetic fields of sunspots forming active regions, shows larger amplitude solar cycle variations than the TSI, though with greater observational uncertainty [23, 32]. This is likely to have a significant effect in the stratosphere [8]. At shorter wavelengths still, episodic solar X-ray flares exhibit a strong solar cycle modulation in occurrence and position across all energy ranges [4]. The Hα flare index is a proxy for the total energy emitted by a flare, expressed as the product of Hα flare intensity and duration in minutes [66], while the X-ray flare index is based on the soft X-ray flux in the 1-8 angstrom range measured by the GOES satellites [16]. Perhaps unsurprisingly, the other major episodic form of solar activity, coronal mass ejections, show a similar occurrence and position variation to flares and sunspots, though there is not one-to-one agreement [29]. See also [31]. Strong solar cycle trends are also present in F10.7 radio flux (10.7 cm or 2.8 GHz), which is a measure of activity in the high chromosphere/low

<table>
<thead>
<tr>
<th>Year</th>
<th>TSI (27-day average)</th>
<th>TSI (1-yr average)</th>
<th>Monthly SSN</th>
</tr>
</thead>
<tbody>
<tr>
<td>1980</td>
<td>1365</td>
<td>1365</td>
<td>1365.5</td>
</tr>
<tr>
<td>1985</td>
<td>1365.5</td>
<td>1365</td>
<td>1366</td>
</tr>
<tr>
<td>1990</td>
<td>1366</td>
<td>1366</td>
<td>1366.5</td>
</tr>
<tr>
<td>1995</td>
<td>1366</td>
<td>1366</td>
<td>1367</td>
</tr>
<tr>
<td>2000</td>
<td>1367</td>
<td>1367</td>
<td>1367</td>
</tr>
<tr>
<td>2005</td>
<td>1367</td>
<td>1367</td>
<td>1367</td>
</tr>
<tr>
<td>2010</td>
<td>1367</td>
<td>1367</td>
<td>1367</td>
</tr>
<tr>
<td>2015</td>
<td>1367</td>
<td>1367</td>
<td>1367</td>
</tr>
</tbody>
</table>

Figure 11: Total solar irradiance (TSI) variations over the period 1978–2013 from the PMOD composite [26] (see also [18]). White and red lines show 27-day and 1-year average values, respectively. The solar cycle is indicated by the black-shaded area, which shows sunspot number, arbitrarily scaled.
corona [83] and is the most continuous and weather-independent measure of solar activity, extending back to 1947. The F10.7 data series is the most widely used solar activity index after the sunspot number.

4 Beyond sunspots: The solar cycle in the heliosphere

In the heliosphere, the solar wind and heliospheric magnetic field (HMF) can be directly measured by spacecraft in situ, though only limited spatial sampling is possible at any given time. In near-Earth space, the solar wind has been regularly (if not continuously) sampled since the early 1960s [59, 43]. Figure 12 shows how the HMF intensity and estimated open solar flux (OSF) varies over the solar cycle. The rise in OSF, and hence HMF intensity, over the solar cycle is the result of the increased rate at which CMEs carry new magnetic flux out into the heliosphere [51, 62]. These HMF observations are representative of global heliospheric variations.

The bottom panel of Figure 12 shows the near-Earth solar wind speed, $V$. At the annual timescale, there is very little variability in near-Earth $V$. Slow solar wind is prevalent throughout the solar cycle, with an increasing (though still minor) contribution from fast wind during the declining phase of the solar cycle. These properties, however, are specific to near-Earth space and inherently local. Globally, the solar wind speed shows much enhanced variation. Also shown in Figure 12 is the global average solar wind speed estimated from extrapolation of photospheric magnetic field observations [71]. At solar minimum, the heliosphere is dominated by fast wind, with slow wind confined to a narrow band around the ecliptic plane. At solar maximum, slow wind extends to all latitudes, resulting in lower global solar wind speed. These model results were verified by the high-latitude Ulysses spacecraft observations [54]. See [61] for more detail. The change in global solar wind momentum and pressure balance with the interstellar medium means that the solar cycle extends even to the size of the heliosphere itself [69].

Near-Earth solar wind conditions control the level of geomagnetic disturbance, primarily through modulation of magnetic reconnection at the day-side magnetopause [19]. Thus the variation in near-Earth HMF intensity is expected to result in a corresponding solar cycle variation in geomagnetic activity. In practice, however, this relation is complex, as the out-of-ecliptic HMF component ($B_Z$, which plays the leading role in determining the geoeffectiveness of the solar wind), is far more stochastic in nature than the HMF intensity [49]. The top panel of Figure 13 shows 27-day and 1-year mean values of $aa_H$ [52], a long-running 3-hourly index which measures the global level of disturbance of the Earth’s magnetic field using one northern and one southern hemisphere geomagnetic station. For annual mean values of $aa_H$, there is a clear solar cycle variation, including the longer-term trends of a rise through the early 20th century followed by a decline from 1980 onwards. Even at the annual resolution,
Figure 12: Near-Earth solar wind variations over the period 1963–2018. White and red lines show 27-day and 1-year average values, respectively. The solar cycle is indicated by the black-shaded area, which shows sunspot number, arbitrarily scaled. Top: Open solar flux. Middle: Heliospheric magnetic field intensity. Bottom: Solar wind speed. The cyan line shows the global mean solar wind speed estimated from extrapolation of photospheric magnetic field observations [71].
however, there are times when geomagnetic activity deviates significantly from the solar cycle (e.g., the peak in $aa_H$ in the solar minimum of 1974 and the minimum in $aa_H$ in the solar maximum of 1980).

To further look at the solar cycle variation in geomagnetic activity, it is useful to define discrete periods of geomagnetic storms using simple $aa_H$ threshold values. In the 3-hourly $aa_H$ dataset, thresholds of 56 nT, 98 nT and 224 nT select the top 5%, 1% and 0.1% of all intervals, respectively. The occurrence of such storms is shown in the bottom panel of Figure 13, as the number of days per year. The moderate (top 5%) – and to a lesser extent, severe (top 1%) – storms do show a close relation to the annual mean $aa_H$ value and hence a strong solar cycle variation. The extreme storms (top 0.1%), however, are not as well-behaved over the solar cycle, though occurrence is generally constrained within the envelope of the solar cycle.

5 Cycle-to-cycle variations

In addition to the Schwabe cycle, the 400-year sunspot record shows the existence of centennial-scale variations in solar magnetic activity. These are corroborated by other solar magnetic proxies, discussed here.

Solar wind conditions prior to the advent of *in-situ* spacecraft observations in the 1960s must be reconstructed from proxy data. As shown in the bottom panel of Figure 14, historic records of geomagnetic field variability allow extremely accurate reconstruction of annual-mean near-Earth HMF intensity and solar wind speed [81, 48]. Unfortunately, such data are not sufficient to reach periods of interest such as the Dalton and Maunder minima. To extend back past 1845, sunspot records can be used to reconstruct global solar wind conditions, such as OSF [76], from which near-Earth conditions can then be obtained [65].

Over the solar cycle, variations in the HMF strength and configuration lead to changes in the modulation of galactic cosmic rays (GCR) from outside the solar system. The resulting variation in GCR flux reaching Earth has been measured with ground-based neutron monitors for more than 60 years [75], and for approximately 25 years prior with ionisation chambers [24]. But the interaction of GCRs with the atmosphere produces cosmogenic radionuclides which are stored in natural reservoirs and hence contain information about heliospheric conditions prior to the advent of instrumental measurements. In particular, $^{14}$C in tree trunks and $^{10}$Be in ice sheets have enabled reconstruction of OSF back nearly ten thousand years [86], albeit typically at sub-decadal time scales, owing to long time constants associated with the terrestrial climate system. (Annual resolution reconstructions are possible back to approximately 1400 A.D. using $^{10}$Be, though the signal to noise increases [55]) As shown in the bottom panel of Figure 14, there is good agreement with the long-term trends in HMF estimated by other methods, though individual solar cycles cannot be resolved. The top panel shows that periods of grand maxima and minima are common with quasi-periodicities of the order 100-300 years.
Figure 13: Geomagnetic activity over the last 150 years. The top panel shows the 27-day (white) and annual (red) mean value of the corrected $aa_H$ index, which measures global geomagnetic disturbance. The black shaded background shows sunspot number, for context. The bottom panel shows the annual occurrence of geomagnetic storms of different magnitudes, with red, blue and white indicating moderate (top 5% of all $aa_H$ values), severe (top 1%) and extreme (top 0.1%), respectively. Note the logarithmic scale and that zero occurrence has been set to 0.1 for plotting purposes.
Figure 14: A summary of the long-term variations in total heliospheric magnetic flux (HMF). Top: Millennial-scale reconstructions of total HMF from cosmogenic radionuclide data [96], scaled to match values in the bottom panel. Bottom: Sunspot (red), geomagnetic (blue) and spacecraft (pink) estimates of total HMF.
6 Future directions and goals

The Sun is a variable G2V star. Like other solar-type stars, its variability in the visible band is of relatively low amplitude. Solar activity parameters appear to place our star at a boundary between stars with and without activity cycles, i.e. between magnetically active and quiet stars. Like on the Sun, the broad-band variability of solar-type stars correlates with the level of chromospheric emission (when normalised by the total bolometric emission of the star). The cycle length on solar-type stars shows correlation with their differential rotation, which is not unexpected given the importance of the $\Omega$ effect in the dynamo process. Cycle-to-cycle variations and multiple magnetic cycles seen on the Sun are also found on other solar-type stars [27].

The importance of understanding stellar activity has dramatically increased with the discovery of an ever-increasing number of exoplanets, as their habitability may be strongly influenced by the magnetic activity of their host stars. As the Sun and its multiple magnetic activity cycles are the longest monitored and studied in the greatest detail, they serve as key sources of information when interpreting shorter-term stellar observations. On the other hand, other stars have a wide range of masses, sizes, surface temperatures, evolutionary states, rotation rates, differential rotation profiles coupled with different levels of magnetic activity. Analysis of stellar characteristics relevant to dynamo-driven magnetic activity can inform and advance solar dynamo studies leading to a deeper understanding of the underlying causes of the solar cycle.

Present solar cycle research is directed both toward the past and future. A key direction is to further uncover and digitize records of past activity. There is an ongoing coordinated effort by the solar cycle community to ever-improve the time series of the fundamentally important sunspot number which covers more than 400 years, since earlier data suffer from temporal gaps and variable quality. The efforts to fill data gaps and make the long-term record more even, are greatly improving our knowledge of long-term solar variability [57]. Furthermore, maintaining the time series of other solar cycle parameters (e.g. the fundamentally important magnetic field, 10.7 cm radio flux, EUV and X-ray flux, eruptive activity, heliospheric in-situ data, etc.), is of vital importance for the long-term continuity of solar cycle research, which is key to our understanding of space weather and helps our advanced technology-dependent society to live in the vicinity of a variable star.

References


28


[55] K. G. McCracken and J. Beer. The Annual Cosmic-radiation Intensities 1391-2014; the annual Heliospheric Magnetic Field Strengths 1391-


