Living Rev. Solar Phys., **10**, (2013), 4
http://www.livingreviews.org/lrsp-2013-4
doi:10.12942/lrsp-2013-4

LIVING REVIEWS
in solar physics

Reconstruction and Prediction of Variations in the Open Solar Magnetic Flux and Interplanetary Conditions

Mike Lockwood
University of Reading
Department of Meteorology
Earley Gate, P.O. Box 243
Reading, Berkshire, RG6 6BB, UK
and
Rutherford Appleton Laboratory
RAL Space
Chilton, Didcot
Oxfordshire, OX11 0QX, UK
email: m.lockwood@reading.ac.uk
http://www.eiscat.rl.ac.uk/Members/mike/

Accepted: 19 July 2013
Published: 9 September 2013

Abstract

Historic geomagnetic activity observations have been used to reveal centennial variations in the open solar flux and the near-Earth heliospheric conditions (the interplanetary magnetic field and the solar wind speed). The various methods are in very good agreement for the past 135 years when there were sufficient reliable magnetic observatories in operation to eliminate problems due to site-specific errors and calibration drifts. This review underlines the physical principles that allow these reconstructions to be made, as well as the details of the various algorithms employed and the results obtained. Discussion is included of: the importance of the averaging timescale; the key differences between "range" and "interdiurnal variability" geomagnetic data; the need to distinguish source field sector structure from heliospherically-imposed field structure; the importance of ensuring that regressions used are statistically robust; and uncertainty analysis. The reconstructions are exceedingly useful as they provide calibration between the in-situ spacecraft measurements from the past five decades and the millennial records of heliospheric behaviour deduced from measured abundances of cosmogenic radionuclides found in terrestrial reservoirs. Continuity of open solar flux, using sunspot number to quantify the emergence rate, is the basis of a number of models that have been very successful in reproducing the variation derived from geomagnetic activity. These models allow us to extend the reconstructions back to before the development of the magnetometer and to cover the Maunder minimum. Allied to the radionuclide data, the models are revealing much about how the Sun and heliosphere behaved outside of grand solar maxima and are providing a means of predicting how solar activity is likely to evolve now that the recent grand maximum (that had prevailed throughout the space age) has come to an end.

Keywords: Geomagnetic activity, Open solar flux, Interplanetary magnetic field, Solar wind, Cosmic rays, Cosmogenic isotopes, Centennial variations

This review is licensed under a Creative Commons
Attribution-Non-Commercial 3.0 Germany License.
http://creativecommons.org/licenses/by-nc/3.0/de/

Imprint / Terms of Use

Living Reviews in Solar Physics is a peer reviewed open access journal published by the Max Planck Institute for Solar System Research, Max-Planck-Str. 2, 37191 Katlenburg-Lindau, Germany. ISSN 1614-4961.

This review is licensed under a Creative Commons Attribution-Non-Commercial 3.0 Germany License: http://creativecommons.org/licenses/by-nc/3.0/de/. Figures that have been previously published elsewhere may not be reproduced without consent of the original copyright holders.

Because a *Living Reviews* article can evolve over time, we recommend to cite the article as follows:

Mike Lockwood,
"Reconstruction and Prediction of Variations in the Open Solar Magnetic Flux and Interplanetary Conditions",
Living Rev. Solar Phys., **10**, (2013), 4. URL (accessed <date>):
http://www.livingreviews.org/lrsp-2013-4

The date given as <date> then uniquely identifies the version of the article you are referring to.

Article Revisions

Living Reviews supports two ways of keeping its articles up-to-date:

Fast-track revision. A fast-track revision provides the author with the opportunity to add short notices of current research results, trends and developments, or important publications to the article. A fast-track revision is refereed by the responsible subject editor. If an article has undergone a fast-track revision, a summary of changes will be listed here.

Major update. A major update will include substantial changes and additions and is subject to full external refereeing. It is published with a new publication number.

For detailed documentation of an article's evolution, please refer to the history document of the article's online version at http://www.livingreviews.org/lrsp-2013-4.

Contents

1 A Brief History of Geomagnetism **5**
1.1 Early observations . . . 5
1.2 The growth of the global magnetometer network . . . 5
1.3 The space age . . . 6

2 Geomagnetic Indices **7**
2.1 Standard geomagnetic indices . . . 7
2.1.1 The *Dst* index . . . 7
2.1.2 The *AU* and *AL* indices . . . 7
2.1.3 The *aa* index . . . 9
2.1.4 The *Am*, *An*, and *As* indices . . . 9
2.1.5 The *Ap* index . . . 10
2.2 Specialist geomagnetic indices . . . 10
2.2.1 The *u* index . . . 10
2.2.2 The *IDV* index . . . 10
2.2.3 The *IDV(1d)* index . . . 11
2.2.4 The *m* index . . . 12
2.2.5 The *IHV* index . . . 12
2.2.6 *"sigma-H"* indices . . . 12
2.3 Dependencies of the various indices on interplanetary parameters . . . 13

3 The Long-Term Variability of Geomagnetic Activity **15**

4 A Note on the Importance of Understanding the Provenance of Geomagnetic Data **21**

5 Solar Wind Coupling Functions, the Importance of Averaging and Allowance for Data Gaps **22**

6 Differences Between Range and Hourly Mean Geomagnetic Data and the Effect of Solar Wind Speed **27**

7 Open Solar Flux, PFSS, and the Ulysses Result **32**
7.1 Open solar flux . . . 32
7.2 The potential field source surface (PFSS) method . . . 32
7.3 Ulysses observations . . . 33
7.4 Excess flux . . . 34
7.5 The use of the modulus of the radial field . . . 38

8 Regression Techniques **42**

9 Reconstructions **46**
9.1 Results for the near-Earth IMF . . . 46
9.2 Results for the near-Earth solar wind speed . . . 47
9.3 Results for the open solar flux . . . 48
9.4 Discussion and uncertainty analysis . . . 51
9.4.1 Comparison of reconstructions and the concept of "floor" values . . . 51
9.4.2 Analysis of uncertainty . . . 53

10 Comparison with Cosmogenic Isotopes **57**

11 Models of Open Solar Flux Variation **60**

12 The Future **63**
12.1 Solar cycle 24 update 1: 20 April 2013 70
12.2 Solar Cycle 24 Update 2: 1 August 2013 70

13 Concluding Remarks **72**

14 Acknowledgments **73**

References **74**

1 A Brief History of Geomagnetism

1.1 Early observations

The magnetic compass was invented in China sometime around 100 AD and was first described in European texts by Guyot de Provins and Alexander Neckam in 1180. Using it, inclination (the angle between the geomagnetic field vector and the horizontal) was discovered in 1576 by Robert Norman. In 1600, William Gilbert (later personal doctor to Elizabeth I of England) published his book "De Magnete" (the full title translates from the latin as "On Loadstone, Magnetic Bodies, and on the Great Magnet of the Earth"), in which he noted that the variation of inclination with latitude was the same as that around a sphere of loadstone and deduced for the first time that Earth had a magnetic field. From observations of the field declination (the angle between the local geographic and geomagnetic northward directions) in London, Henry Gellibrand concluded in 1634 that the geomagnetic field changes over time (i.e., he had discovered the secular variation) and in 1701 Edmund Halley published the first geomagnetic field map, showing the declination throughout the Atlantic ocean. By 1722, the London watchmaker George Graham had developed a compass that was sufficiently sensitive for him to observe the fluctuations that we now call "geomagnetic activity" (Graham, 1724a,b). The diurnal variation in that activity was first noted in 1740 by Olof Hiorter, Anders Celsius' student (and brother-in-law), in Uppsala. On the night of 1 March, 1741 Hiorter also observed large magnetic variations that were connected to local auroral displays and, on the same night, magnetic variations were also recorded in London by Graham, making these the first multi-point observations of geomagnetic activity.

> *"Who could have thought that the northern lights would have a connection and a sympathy with a magnet?"*
> – Olof Peter Hiorter, Swedish Scientist (1696 – 1750)

The relative intensity of the geomagnetic field at a number of locations was measured in 1798 by Alexander von Humboldt, who was the first to describe the disturbances he observed as "geomagnetic storms". His work attracted the interest of his house guest, Carl Friedrich Gauss, when visiting Berlin to attend a conference in 1828. Gauss subsequently developed the first magnetometer that could reliably measure the field strength and/or its horizontal component, establishing the first magnetic observatory in Göttingen in 1832. He also established the mathematics of how to separate the internal and external components of Earth's magnetic field.

1.2 The growth of the global magnetometer network

The year after the establishment of the Göttingen observatory, Gauss and Wilhelm Weber founded the "Magnetischer Verein" (Magnetic Union) which from 1834 to 1841 initiated the growth of a network of observatories throughout Europe making measurements at 5-minute intervals. Magnetometers were soon established at sites such as Berlin (1836), Dublin (1838), Greenwich (1838), Prague (1839), and Munich (1841). Gauss and Weber began to organize a global magnetic survey, an idea strongly supported by the scientific adviser to the British Admiralty, Edward Sabine. As a result, the British Navy set up more stations in Toronto, St. Helena, Cape of Good Hope, and Tasmania and the British East India Company established four more in India and Singapore. Russia established ten stations in its own territory (which at the time included Helsinki) and one in Beijing, and by 1841 a world-wide network of 53 stations was operating. Using data from these observatories, Sabine was the first to realize that geomagnetic activity could be divided into a regular diurnal cycle and irregular variations which correlated very closely with sunspot number (Sabine, 1851, 1852), the decadal-scale cycle of which had been discovered in 1844 by the German astronomer Heinrich Schwabe. The subsequent development of solar-terrestrial science using the

geomagnetic activity observations has been detailed in three excellent reviews by Cliver (1994a,b, 1995).

The number of available magnetic observatories subsequently grew gradually over the next century, helped by international campaigns such as the Polar Year (1882 – 1883) and the Second International Polar Year (1932 – 1933), so that by 1955 about 100 stations worldwide were supplying regular routine observations. This number rose rapidly because of the International Geophysical Year, IGY (1957 – 1958), reaching of order 170 by 1960 (Jankowski and Sucksdorff, 1996). Figure 1 shows the global distribution of stations known to be operating in 1996.

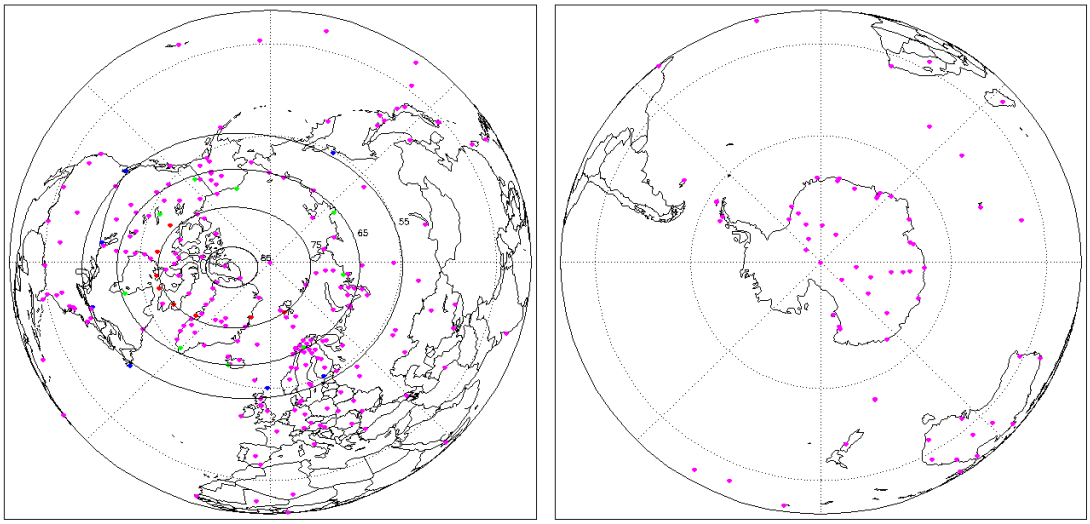

Figure 1: Pole-centric maps of the (left) northern and (right) southern hemispheres showing the locations of known operating magnetometer stations in 1996 (coloured dots), as compiled by the World Data Centre, C1 for Solar Terrestrial Physics, RAL Space, UK.

1.3 The space age

Modern understanding of geomagnetic activity relies heavily on in-situ spacecraft observations of the solar wind, shortly before it impacts on the Earth. Such measurements were first made routinely in 1963 but the monitoring was not close to continuous until about 1966. After a few years the length and number of data gaps in this vital space science resource began to increase (Finch and Lockwood, 2007), driven by factors such as telemetry limitations and a shortage of available tracking stations. In this respect, 1995 is a significant date in that (almost completely) continuous solar wind monitoring began with the WIND spacecraft and has continued with the ACE spacecraft to the present day. Covering almost two solar cycles, these continuous data constitute the most valuable resource we have for understanding how solar wind properties, including its bulk flow speed, V_{SW}, and the interplanetary magnetic field (IMF) embedded within it, B, drive geomagnetic activity. The near-Earth interplanetary data have been collected by NASA's Goddard Space Flight Centre (the Space Physics Data Facility) into the OMNI and OMNI2 datasets (Couzens and King, 1986; King and Papitashvili, 2005).

2 Geomagnetic Indices

A large number of indices have been developed and deployed to quantify the geomagnetic activity detected by the global network of magnetic observatories. These indices vary in which observatories are used, which data from those observatories are used, how those data are processed, and how the data from different observatories are combined together. As a result, the different indices monitor different parts of the system of coupled currents that flow in near-Earth space in response to the flow of the magnetised solar wind plasma around the magnetosphere. Figure 2 is a schematic showing the currents that flow in the magnetosphere-ionosphere system and how they are connected.

The magnetic field at a point is the summed effect of all moving charges particles in the cosmos on that point. Because the Biot–Savart law contains an inverse-square dependence on the distance between the moving charges and the point in question, the effects of closer currents tend to dominate over more distant ones but all contribute. As a result, although the deflections seen by a ground based magnetometer usually reflect changes in the closer large-scale currents in the magnetosphere-ionosphere system, there will also always be some effects of other currents flowing elsewhere. The following subsections briefly outline indices that will be employed in this review. In some relatively clear-cut cases, such as *Dst*, *AU*, and *AL*, there is discussion of the currents in near-Earth space which contribute most to the detected variations in the index. However, for other cases the combination of currents that the index is monitoring is not so straightforward, as will be discussed in Section 6. Section 2.1 lists standard indices in widespread use whereas Section 2.2 discusses some research indices, designed for reconstruction work using historic datasets. Section 2.3 presents an initial study of how these various indices vary with parameters describing near-Earth interplanetary space.

2.1 Standard geomagnetic indices

2.1.1 The *Dst* index

The *Dst* (Disturbed Storm Time) index is constructed using hourly means of the horizontal component H measured at four equatorial magnetometer stations: Honolulu, San Juan, Hermanus, and Kakioka. The index was first constructed for the International Geophysical Year and is available for 1957 onwards. The derivation and station selection is described by Sugiura and Kamei (1991). Because of the low latitudes of the stations, the *Dst* index chiefly monitors the disturbances produced by changes in the ring current, which flows westward around the magnetosphere at geocentric distances of about $3-6\,R_E$ (where $1\,R_E$ is a mean Earth radius), as shown in part (b) of Figure 2. Negative perturbations in Dst correspond to storm time enhancements in the ring current. However there are also small contributions from the cross-tail sheet current in the magnetotail and some contamination from auroral ionospheric currents. In addition, positive variations in *Dst* are caused by the compression of the magnetosphere due to solar wind dynamic pressure increases, showing that it also responds to changes in the magnetopause (Chapman–Ferraro) currents.

2.1.2 The *AU* and *AL* indices

The Auroral Electrojet indices (*AE*, *AL*, *AU*, and *AO*) were first introduced by Davis and Sugiura (1966) to a measure the auroral electrojet currents that flow in the high-latitude ionosphere. In order to achieve this, the stations contributing to the index lie within the band of the auroral oval. A ring of 12 longitudinally-spaced magnetometers ensures that one station is always close to the peak of the westward auroral electrojet whilst another station is always close to the peak of the eastward electrojet (Tomita *et al.*, 2011). The stations are all in the northern hemisphere and a corresponding southern hemisphere ring is precluded by the southern oceans which do not allow sufficiently even and full longitudinal coverage of the southern auroral oval. The exact number

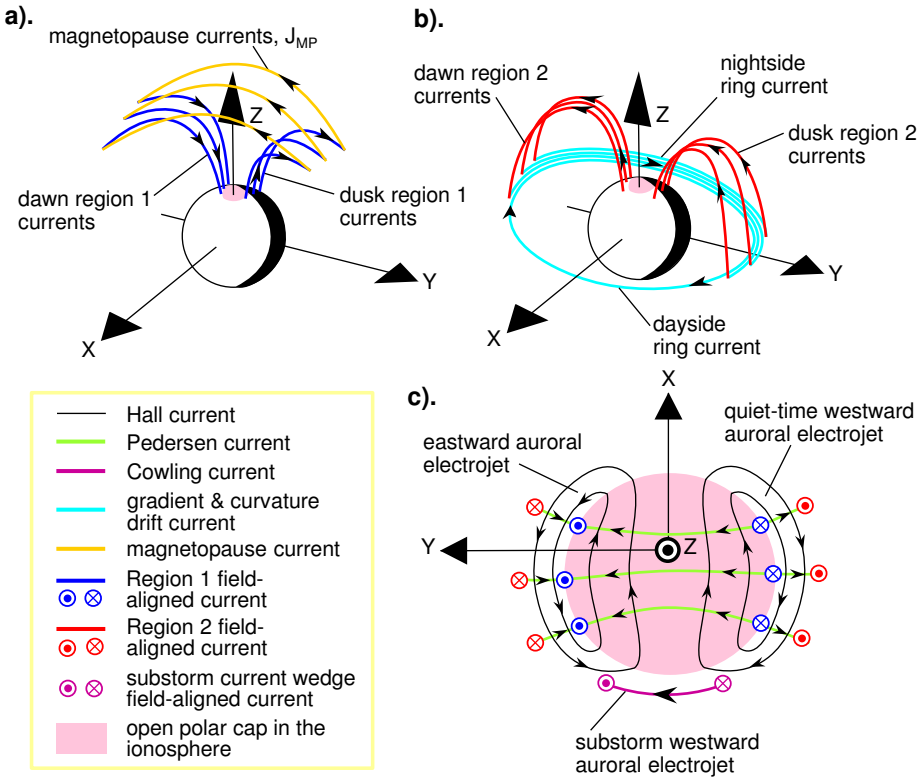

Figure 2: Simplified schematic of the currents flowing in the magnetosphere-ionosphere system in the northern hemisphere (southern hemisphere currents are omitted for clarity). Part (a) shows (in orange) segments of the Chapman–Ferraro currents that flow in the magnetopause and separate the geomagnetic and (shocked) interplanetary fields: the relevant segments are at the sunward edge of the magnetospheric tail and flow from dusk to dawn (see also Figure 14). These connect to the high-latitude ionosphere via the Region 1 field-aligned (Birkeland) currents (shown in blue). The Region 2 field-aligned currents are needed to maintain ionospheric current continuity and because of the incompressibility of the ionosphere (in the sense that the magnetic field there is essentially constant). As shown in red in (b), these Region 2 currents close via the ring current that flows westward around the Earth in the inner magnetosphere (in cyan), caused by the gradient and curvature drifts of trapped energetic particles. Part (c) shows the Region 1 and 2 currents entering and leaving the polar E-region ionosphere and how they connect to the Pedersen currents there (in green), which flow in the direction of the electric field. The paired up and down field-aligned currents transfer solar wind energy, momentum, and electric field down into the ionosphere as well as current (see review by Lockwood, 1997). The Hall currents (shown by black lines) flow perpendicular to the electric field (and so cause no energy dissipation) and are antiparallel to associated ionospheric flow (convection) in the over-lying F-region ionosphere. For a uniform spatial distribution of conductivities, the effects of field-aligned and Pederson currents cancel beneath the ionosphere and only the Hall currents are detected by high-latitude magnetometers on the ground. The formation of the westward electrojet in the substorm current wedge (shown here in mauve) is described later by Figure 14. In this electrojet, a highly conducting channel is formed by ionisation generated by the associated particle precipitation and the Cowling conductivity is relevant.

of stations has varied somewhat over time, data prior to 1964 coming from a somewhat different distribution of stations including contributions from the southern hemisphere. The indices can be generated using a great many stations, but the standard indices employ 12 and are referred to as $AE(12)$. They are recorded at high time resolution (usually 2.5 minutes) and quiet time diurnal variations are first removed from the H component. The maximum and minimum values of the background-subtracted H at any one time seen by the ring stations are the AU and AL values, respectively. (Also often quoted are AE, the difference between AU and AL, and AO, which is their mean, but neither are used in this review). The auroral currents causing geomagnetic activity are divided into the "DP1" and "DP2" systems (e.g., Clauer and Kamide, 1985). Studies of the station contributing the maximum deflection (e.g., Tomita *et al.*, 2011) reveal that large (negative) perturbations to AL are caused by the nightside westward electrojet (the DP1 or substorm current wedge system, see Section 6) which responds to magnetic energy that is stored in the magnetotail and then explosively released into the westward auroral electrojet during events called "substorm expansion phases", whereas AU is set by the eastward part of the dayside DP2 currents that are directly driven by the solar wind (e.g., Clauer and Kamide, 1985; Consolini and De Michelis, 2005). Under quiet conditions AL reflects the westward electrojet of the DP2 system in the morning sector (DP2 currents are generally detected on the dayside where ionospheric conductivities are higher). The eastward, quiet westward and disturbed westward auroral electrojets are all labelled in part (c) of Figure 2. There is some contamination of AU and AL from the ring current. Data are available for 1957 onwards.

2.1.3 The *aa* index

The aa index was devised and compiled by Mayaud (1971, 1972, 1980). It is a "range" index, meaning it is based on the range of variation seen during three-hour intervals, as introduced by Bartels *et al.* (1939). At each station contributing to the index, a semi-logarithmic k index is derived by first removing the quiet-time variation and then using the larger of the differences between the maximum and minimum values of either the horizontal or vertical field in the 3-hourly intervals (the range), giving eight values per day. Data are taken from just two mid-latitude stations selected to be close to antipodal, with the northern hemisphere station in southern England and the southern in Australia. In both hemispheres, three different stations were needed to give a continuous index: in the north they are Greenwich (1868 – 1925), Abinger (1926 – 1956), and Hartland (1957 – present) and in the south they are Melbourne (1868 – 1919), Toolangi (1920 – 1979), and Canberra (1980 – present). The k indices are generated using a site-dependent scale to normalise them to the values seen at the Niemegk station, giving aa_N and aa_S for the north and south hemispheres and aa is defined as the arithmetic mean of the two.

2.1.4 The *Am*, *An*, and *As* indices

The Am, An, and As indices are range indices constructed in the same way as aa but use a greater number of stations. Mid-latitude stations (around a target geomagnetic latitude of 50°), spread across geomagnetic longitudes in both the northern and southern hemispheres, are used. The exact mix of stations has varied somewhat over time, as have the longitudinal sectors into which they have been divided, but typically number 16 in the north and 9 in the south. The k values are averaged over the longitudinal sectors (5 in the north, 4 in the south) before being normalised and then averaged over the northern hemisphere, southern hemisphere and globally to give An, As, and Am, respectively. Data are available from 1959.

2.1.5 The *Ap* index

The *Ap* index is another range index, which is available for 1932 onwards. It is a 3-hourly planetary index compiled using the k indices from 11 – 13 longitudinally-spaced mid-latitude stations in the northern hemisphere.

2.2 Specialist geomagnetic indices

The *Dst*, *AU*, *AL*, *Am*, *An*, *As*, and *Ap* indices are all well-established, in widespread use and formally recognised by international organisations such as IAGA (the International Association of Geomagnetism and Aeronomy). However there are other valuable indices that have been compiled by individual researchers to meet specific purposes. These generally employ hourly mean or hourly "spot values" (samples).

2.2.1 The *u* index

The u index was developed by Bartels (1932). It was based on the absolute value of the difference between the mean values of H for a day and for the preceding day. Taking this difference is a simple but effective way of removing quiet time variation. The u index is the weighted mean of data from a collection of stations. Prior to averaging the data from the various stations, each was normalised to the magnetic latitude (Λ) of Niemegk using an empirical $1/\cos(\Lambda)$ dependence. Bartels used data from Seddin (1905 – 1928), Potsdam (1891 – 1904), Greenwich (1872 – 1890), Bombay (1872 – 1920), Batavia (1884 – 1899 and 1902 – 1926), Honolulu (1902 – 1930), Puerto Rico (1902 – 1916), Tucson (1917 – 1930), and Watheroo (1919 – 1930). He notes stability problems with the Greenwich data in deriving interdiurnal variation data (from one day to the next) and ascribes half weighting to it as a result. (Recently, Lockwood *et al.*, 2013a have studied all the hourly data from Greenwich and confirmed these problems). In addition Bartels notes many data gaps in the Bombay data. The u index is based on 2 stations for 1872 – 1891 (Greenwich and Bombay and Bartels expresses reservations about the quality of both), rising to 6 by 1919 before falling to 3 again by 1930. Data for 1835 – 1872 was compiled by Bartels and is called the u index but is not the same as the index after 1872. Bartels notes that before 1872, no proper data to generate an interdiurnal index was available to him and so other correlated measures of the diurnal variation are used as proxies. Bartels himself stresses that the u values before 1872 are "more for illustration than for actual use". The u index was criticised at the time for failing to register the recurrent geomagnetic storms and, as a result, he himself developed the range indices as an alternative (Bartels *et al.*, 1939). However, as pointed out by Svalgaard and Cliver (2005), this feature is a positive advantage of u as it means that it is not complicated by a response to solar wind speed variations. The u index data cease in 1930.

2.2.2 The *IDV* index

The IDV index is a variant of the u index that was devised by Svalgaard and Cliver (2005). The main difference in its derivation is that instead of using daily mean values of H, the hourly mean (or spot value) closest to solar local midnight is employed. As for u, the difference between values on successive days is taken. The latitude normalisation is also slightly different, using an empirically-derived $1/\cos^{0.7}(\Lambda)$ dependence. IDV is found to not depend on solar wind speed, V_{SW} and depends on just the IMF field strength B in annual means. One of the great advantages of IDV is that it's compilation is much simpler than the range-based indices and this has allowed the use of historic hourly mean (or spot) values to produce a meaningful index that extends back many years. Svalgaard and Cliver (2005) adopt a different philosophy in compiling IDV to that adopted by Mayaud (1971) in compiling *aa*. Mayaud's philosophy was to use as homogeneous a

data series as possible. The philosophy of Svalgaard and Cliver (2005) (and of Lockwood *et al.*, 2006b in the derivation of the m index, see below) was to use all available data that are of sufficient quality. Inevitably this means that fewer data are available at earlier times and the construction of IDV means that (like m) it is not homogeneous. Svalgaard and Cliver (2010) added more stations and also extended the sequence back to 1835 using a linear correlation with the Bartels u index. In this context, note Bartels' reservations about the early u data discussed in Section 2.2.1.

2.2.3 The *IDV(1d)* index

The $IDV(1d)$ index has recently been introduced by Lockwood *et al.* (2013a). This is very similar to IDV with two differences. The first is that it employs daily means rather than the near-midnight hourly mean or spot values: in other words, Lockwood *et al.* (2013a) returned to the formulation used by Bartels (1932) to generate u. This means that 24 times the volume of data are used than in generating IDV because data from the 23 UT-hours away from local midnight are not discarded. This has advantages in noise suppression by averaging. Lockwood *et al.* (2013a) adopted the name $IDV(1d)$ (rather than reverting to the name u) because of the other major difference, namely that the $IDV(1d)$ composite is homogeneous in its construction (i.e., it uses the philosophy of aa and not that of u and IDV), using data from three intercalibrated stations sequentially to form a composite. Data from Helsinki were used for 1846 – 1890 (inclusive) and 1893 – 1897 and from Eskdalemuir from 1911 to the present day. The gaps are filled using data from the Potsdam (1891 – 1892 and 1898 – 1907) and the nearby Seddin observatories (1908 – 1910) and intercalibration achieved using the Potsdam/Seddin/Niemegk data sequence for 1890 – 1931. To remove site effects and the effects of secular drifts in geomagnetic latitudes, the $1/\cos^{0.7}(\Lambda)$ dependence found by Svalgaard and Cliver (2005) was shown to apply and was used to make a small ($> 5\%$ between 1846 and 2013) correction to the data based on model predictions of the magnetic latitude of the stations, Λ. The $IDV(1d)$ index extends back to the start of the Helsinki data in mid 1845. One key justification of $IDV(1d)$ is that it correlates with the IMF B as well as (in fact very slightly better than) IDV, despite the fact that it is based on data from just one station (which is Eskdalemuir throughout the space age) rather than the approximately 50 stations contributing to IDV at that time (see Figure 3). One concern, however, is the use of the historic data from Helsinki which is at a higher corrected magnetic latitude and so more subject to auroral current contamination of the kind noted by Svalgaard and Cliver (2010) and Finch *et al.* (2008), and which could introduce a dependence on solar wind speed, V_{SW}. Models of the geomagnetic field give corrected magnetic latitudes of Helsinki varying between 55.5° and 56.5° over the interval that data are used from this station. The survey by Finch et al found that the correlation with IMF B began to drop above 60° (and that with V_{SW} began to rise). Hence, Helsinki is close to being at too high magnetic latitude. To investigate if this was a problem, Lockwood *et al.* (2013a) used modern data from the Nurmijärvi station (close to Helsinki) and compared IDV(1d) derived from them to that from Eskdalemuir. The correlation is 0.931 in 27-day means and 0.982 in annual means. Furthermore, the dependence of $IDV(1d)$ from Nurmijärvi on BV_{SW}^n was investigated and the peak correlation found near $n = 0$, very close to the value for Eskdalemuir (see Figure 3). The same tests were applied to modern data from the Niemegk station.

The homogeneous nature of $IDV(1d)$ is a major advantage when making historic reconstructions of interplanetary parameters because one can have greater confidence that it will have responded to changes in the solar wind before the space age in the same way that it was observed to do during the space age. If an index is not constructed in a homogeneous manner then one cannot have that confidence to the same extent. Hence, for reconstructions of interplanetary parameters, homogeneously constructed indices such as $IDV(1d)$ and aa are preferable to inhomogeneous ones such as IDV and m.

2.2.4 The *m* index

The m index was introduced by Lockwood *et al.* (2006b) and used by Rouillard *et al.* (2007) and Lockwood *et al.* (2009d). For each station at a given UT, the standard deviation of the hourly means of the horizontal component of the geomagnetic field is computed over a full year, $\sigma^H_{1\,\mathrm{yr}}$. These are then correlated with, and linearly regressed against, the annual means of aa_C shown in Figure 5 to yield $\sigma' = s \times \sigma^H_{1\,\mathrm{yr}} + c$. These normalisations are needed because both the sensitivity s and offset c for a station were shown to depend on its location and on the UT hour (which, for example, alters the location of the station relative to the midnight-sector auroral oval) (Finch, 2008). Each station-UT is treated as an independent data series. The median of these data series is used as it is less influenced by extreme outliers than the arithmetic mean. Somewhat conservative criteria are used for the inclusion of data, in that annual means of the station-UT time series must correlate reasonably well (correlation coefficient > 0.5) with those of aa_C. In addition, m does not employ any isolated fragments of data from stations that ceased operating before the start of the space age and only used data from stations that continued to take data into the space age (or there was a nearby station, with which one could make a composite, that did). The advantage of m over IDV is that data from all 24 UT-hours are employed, as opposed to just the one (near midnight) value used by IDV. The disadvantage is that its compilation is much more complex and time-consuming than that of IDV and so new data cannot be as readily added. Furthermore, m does not correlate as highly with interplanetary parameters as does IDV (see Figure 3). Lockwood *et al.* (2006b) consider that the m index is less reliable before 1902 because then it is based on data from just one station (Potsdam).

2.2.5 The *IHV* index

The IHV index was devised and introduced by Svalgaard *et al.* (2003) and Svalgaard and Cliver (2007a) and uses only nightside data to minimise the effect of the diurnal variation. IHV for a given station is defined as the sum of the absolute values of the difference between hourly means (or spot values) for a specified geomagnetic component from one hour to the next over the 7-hour interval around local midnight. The variation with the corrected magnetic latitude shows strong peaks in the auroral oval, indicating it responds most to the variability in the nightside westward auroral electrojet and so it behaves rather like AL. Because the variation with corrected geomagnetic latitude is flat equatorward of 55° only stations equatorward of this were employed in the global IHV index. The normalisation, grouping and averaging of data from different stations to obtain a global index is described in Svalgaard and Cliver (2007a).

2.2.6 *"sigma-H"* indices

Finch *et al.* (2008) devised the σ^H_n indices for each station, which – like IDV, $IDV(1d)$ and m – is based on hourly mean data. σ^H_n at a given station is defined as the value of the standard deviation of the hourly-averaged H values at a given UT over a period of n days, each single UT-hour being treated separately, as for m. There will therefore be 24 values for each period of n days at each station. (Note that the m index is, using this notation, the median for all available station-UTs of the σ^H_{366} values for leap years and σ^H_{365} for all other years). Finch *et al.* (2008) used $n = 28$ days (close to the solar rotation period, as seen from Earth, which is the Carrington rotation period of 27.2753 days), which gives thirteen 28-day periods per year (with any excess days assigned to the final such period in the year). It is a different measure of the inter-diurnal variation quantified by the u, IDV, $IDV(1d)$ indices, but because it applies to each station individually it is, in some respects, equivalent to the k values used in the derivation of range indices.

2.3 Dependencies of the various indices on interplanetary parameters

In-situ spacecraft data on the near-Earth interplanetary medium became increasingly available from 1963, at the start of the space age. Early studies comparing geomagnetic activity to the near-Earth interplanetary parameters (e.g., Arnoldy, 1971) showed that geomagnetic activity was enhanced when the interplanetary magnetic field (IMF) pointed southward in a reference frame aligned by Earth's magnetic axis: Geocentric Solar Magnetospheric, GSM, is widely used (Russell, 1971; Hapgood, 1992). This had been predicted in the seminal paper by Dungey (1961), who proposed that for this IMF orientation, magnetic reconnection in the dayside magnetopause current sheet would allow the solar wind to drive stronger F-region ionospheric flows (convection) and hence the associated E-region ionospheric currents and geomagnetic activity seen at Earth's surface would also be stronger. The southward IMF orientation in GSM occurs for 50% of the time (Hapgood *et al.*, 1991). The DP2 or "directly driven" currents respond to IMF variations with a lag of a few minutes (Nishida, 1968), whereas the larger DP1 or "storage-release system" currents are enhanced during substorm expansion phases following a lag of typically one hour (e.g., Baker *et al.*, 1981). The high latitude auroral currents link to the magnetospheric ring current via the Region-2 field-aligned currents, as shown in Figure 2. The ring current has long been understood in terms of injection and decay of the trapped particles that carry it (Burton *et al.*, 1975) and the injection is more efficient when the interplanetary magnetic field points southward (see, e.g., Shi *et al.*, 2012). The response is complicated by the fact that the interplanetary electric field also influences the decay of the ring current and there are other, internal magnetospheric factors which influence both the injection and the decay (see reviews by Kozyra and Liemohn, 2003; Pulkkinen, 2007). Enhancements of the ring current cause negative depressions in the Dst index but will also influence other geomagnetic indices.

Figure 3 explores the dependence, on annual averaging timescales, of the geomagnetic indices described on Sections 2.1 and 2.2 on the solar wind speed, $V_{\rm SW}$. The correlation between each index and $BV_{\rm SW}^n$ is presented where B is the IMF field strength and n is an exponent that is here varied between –2 and 4. The correlations are for annual means between 1966 and 2012, inclusive. Parameters marked with a prime denote that data have been omitted in computing both sets of annual means if any of the simultaneous (allowing for the predicted satellite-to-Earth solar wind propagation lag) hourly means of B, $V_{\rm SW}$ or the geomagnetic index are missing due to a data gap. In the case of the 3-hour range indices aa, Am and Ap, the procedure adopted by Finch and Lockwood (2007) is followed to ensure only simultaneous geomagnetic and IMF data are included in the annual means. In the case of $IDV(1d)$, each daily value contains information on H from two whole days: in order to be included in the annual means, we here require that there be 75% coverage of the IMF observations over those two days. The value of 75% is chosen as a compromise between not eliminating too much of the data and removing data for which the interplanetary means could be misleading because the data coverage is low. The effects of not carrying out this piecewise removal of data from both sets during datagaps were studied by Finch and Lockwood (2007): effectively one is assuming that annual means are representative, even when large fractions of the data are missing (as they are in some years for the interplanetary data). Even with the piecewise removal of data during data gaps, we here only employ annual means that have data availability exceeding 50% to avoid years of reduced data having undue weight. In the study presented in Figure 3, all the correlations are somewhat improved by taking these steps and, importantly, the n of peak correlation is sometimes also affected. Note that only annual mean data for IDV and IHV have been published and the way m is generated only yields annual values: as a result, no allowance for gaps in the interplanetary data can be made in these three cases (hence there is no prime symbol attached to IDV, IHV, or m in Figure 3).

The coupling functions $BV_{\rm SW}^n$ have been calculated in hourly data and then averaged, so that $\langle BV_{\rm SW}^n\rangle_{1\,\rm yr}$ is used rather than $\langle B\rangle_{1\,\rm yr}(\langle V_{\rm SW}\rangle_{1\,\rm yr})^n$. The AL auroral electrojet index (red line)

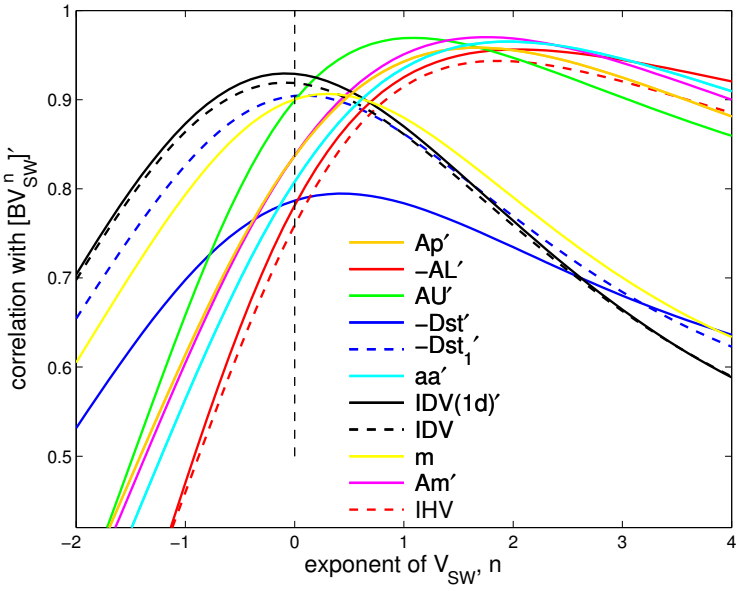

Figure 3: Linear correlation coefficients of annual means of various geomagnetic indices with $BV_{\rm SW}^n$, as a function of n, the exponent of the solar wind speed, $V_{\rm SW}$ (B is the IMF field strength). The primes denote the fact that data have been omitted in calculating either set of annual means if any of $V_{\rm SW}$, B or the geomagnetic index are missing because of data gaps exceeding 1 hour duration. Values of $BV_{\rm SW}^n$ are computed hourly and then averaged. Correlograms are shown for: (red line) AL; (green line) AU; (blue solid line) $-Dst$; (blue dashed line) the negative part of Dst, $-Dst_1$ (where Dst_1 is the same as Dst but intervals when $Dst > 0$ are treated as data gaps); (cyan) aa; (orange) Ap; (black dashed) IDV; (black solid line) $IDV(1d)$; (mauve) Am; (yellow) m and (red dashed) IHV. For indices which are increasingly negative for increasing activity (Dst, Dst_1 and AL) the index has been multiplied by -1. Image reproduced from Lockwood *et al.* (2013a).

shows peak correlation $n = 2$, i.e., it has a $BV_{\rm SW}^2$ dependence. The AU index (green line) gives a peak at $n = 1.1$ (i.e., it has close to a $BV_{\rm SW}$ dependence and, hence, varies with the interplanetary electric field). The Dst index shows a peak at $n = 0.4$ (blue line) but some of this dependence on $V_{\rm SW}$ arises from the compression of the equatorial field by enhanced solar wind dynamic pressure: if we use only the negative part of Dst (Dst_1, which is the same as Dst but treats all intervals where $Dst > 0$ as data gaps and so only contains intervals when Dst is dominated by ring current effects), we get the dashed blue line with a higher correlation coefficient peak at $n = 0.1$. This peak is flat and, hence, the peak n is not significantly different from zero (i.e., the dependence is on B alone). The cyan line is for the aa index and peaks at $n = 1.9$ (very close to the $BV_{\rm SW}^2$ dependence of AL), the mauve line is for Am and peaks at $n = 1.8$ and the orange line is for Ap and peaks at $n = 1.6$. The black line is for the $IDV(1d)$ index, which peaks at n near -0.1. Hence $IDV(1d)$, like IDV and Dst_1, is not significantly different from having a dependence on B only. Thus, as concluded by Svalgaard and Cliver (2010), the negative part of Dst (i.e., ring current enhancement) is closest to explaining the behaviour of the interdiurnal variability indices on these annual timescales. The range indices, aa and Ap respond in a manner similar to the auroral indices and, in particular, the influence of the westward auroral electrojet on aa (as monitored by AL) can be inferred from the fact that both have a dependence that is not significantly different from $BV_{\rm SW}^2$. The correlation for Ap peaks at a slightly lower n than for AL, Am, or aa, which may be a greater influence of the directly-driven currents or may be the effect of the ring current (as both AU and Dst give peaks at lower n). The IHV index correlation peaks at $n = 1.9$ and so, as

pointed out by Svalgaard *et al.* (2003) and Svalgaard and Cliver (2007a), behaves very much like AL and all the range indices with a BV_{SW}^2 dependence, as expected because it is a monitor of the nightside auroral electrojet.

The m index (yellow line) correlation peaks at $n = 0.3$ and there are a number of possible reasons why this value of n exceeds zero. It could be that the response of m is set by a mixture of the ring current (with its $n = 0$ dependence) and the DP2 auroral currents (with their $n = 1$ dependence). An alternative explanation is that the normalisation against the aa index in the derivation of m has introduced a small dependence on V_{SW}. We also note that m employed data from some auroral stations such as Sodankylä, which, as discussed in Section 6, introduces a V_{SW}^2 dependence into σ_H values.

3 The Long-Term Variability of Geomagnetic Activity

The first homogeneous, long-term record of geomagnetic activity was the aa index compiled by Mayaud, who analysed 100 years' data (1868 – 1968) from observatories in southern England and near their antipodal locations in Australia (Mayaud, 1971, 1972, 1980). In each hemisphere, three different stations were required to make a continuous record and, as for all such composites, inter-calibration problems between the different stations arise. Means over calendar years and over 27-day Bartels solar rotation intervals are plotted in the top panel of Figure 4. When aa was first used to reconstruct the solar magnetic fields, there were several vociferous objections that, despite Mayaud's careful calibration work, the drift seen in Figure 4 was merely an instrumental artefact. There are, indeed, a great many problems that can cause long-term changes in the record from a magnetometer station: in addition to instrument changes, drifts and re-adjustments, a change in the local water table can have an influence, as can the construction of power or railway lines nearby and, on very long timescales, the secular drift in the magnetic poles of the Earth causes the geomagnetic coordinates of a station to drift. The argument that the aa values could not be as low as derived around 1900 by Mayaud has been proved to be wrong by the recent long and low solar minimum between solar cycles 23 and 24 (Russell *et al.*, 2010; Lockwood, 2010). During this minimum (around 2008), comparably low annual mean aa was observed, as is shown by Figure 4. Analyses suggesting calibration problems were often based on comparisons with hourly mean data (Svalgaard *et al.*, 2004; Svalgaard and Cliver, 2005; Mursula and Martini, 2007). However, it has become clear that hourly mean data are not observing the same mix of currents and phenomena as the range indices (see Section 6) and, hence, many of the differences are real rather than instrumental. This is underlined by the bottom panel of Figure 4, which shows the corresponding means for another homogeneously-constructed long-term index, $IDV(1d)$. Although many features can be seen in both indices, there are many differences, particularly in the 27-day means. The long-term change is seen in both indices despite 3 major differences between them: (1) they are constructed using data from entirely different observatories, (2) one uses hourly means and the other range data, and (3) the compilation algorithms (including the removal of quite day variations and secular change in station latitudes) are entirely different. The correlation coefficients between aa and $IDV(1d)$ for 27-data and annual means (for 1868 – 2013) are 0.68 and 0.76, respectively. These correlations should be compared with those over the same interval between the independent aa indices for the northern and southern hemispheres, aa_N and aa_S, which are 0.94 for the 27-day means and 0.98 for the annual means.

A number of tests on aa have been carried out (e.g., Lockwood, 2001; Clilverd *et al.*, 2002; Cliver and Ling, 2002; Lockwood, 2003; Clilverd *et al.*, 2005; Lockwood *et al.*, 2006b; Lu *et al.*, 2012) which show it to be a reasonable indicator of long-term change. Furthermore, studies of potential factors identified a solar origin of the long-term drift (Clilverd *et al.*, 1998; Stamper *et al.*, 1999; Clilverd *et al.*, 2002). However, there is also evidence of some error in the aa index, as stored in

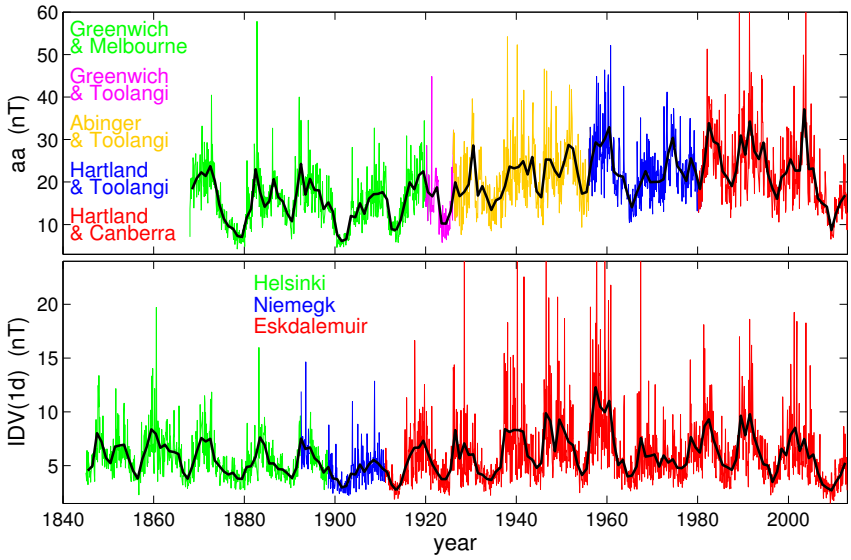

Figure 4: Averages over Bartels 27-day solar rotation intervals (coloured) and calendar years (in black) of (top) the *aa* geomagnetic index and (bottom) the $IDV(1d)$ index. The colours give the station(s) contributing the data to the indices at any one time. This version of the *aa* index is as stored in the World Data Centres. The data labelled Niemegk are a composite from the three nearby stations, Niemegk, Potsdam, and Seddin.

most data centres at the present time. The first authors to suggest errors in *aa* were Svalgaard *et al.* (2004) who compared *aa* against the IHV index: indeed they argued that all centennial change in *aa* was erroneous. Comparing with IHV is a valid test of *aa* because, as shown in Figure 3, they both correlate best with $BV_{\rm SW}^{n}$ for n near 2. The initial comparisons by Svalgaard *et al.* (2004) found an almost negligible change in IHV since 1900 which would imply early *aa* values were too low: quantitatively they found the mean error in *aa* was 8.1 nT over solar cycle 14 (1901 – 1912, inclusive), which considering that the mean *aa* over this cycle was lower than the mean for cycles 20, 21, and 22 by the same amount, means that they argued that all the long-term change in *aa* was erroneous. However, this early version of IHV was based on just one composite data series from two very nearby stations, Cheltenham and Fredricksburg (intercalibrated using the available 0.75 yr of overlapping data in 1956). Using more stations, Mursula *et al.* (2004) found there was upward drift in IHV values over the 20th century, but it depended on the station studied; nevertheless they inferred that the drift in *aa* was too large. As a result, Svalgaard *et al.* (2003) revised their estimates, using several stations, such that the cycle-14 mean of *aa* was too low by 5.2 nT (this would mean that 64% of the drift in *aa* was erroneous). However, Mursula and Martini (2006) showed that about half of this difference was actually in the IHV estimates not *aa* and was caused by the use of spot values rather than hourly means in constructing the early IHV data. This was corrected by Svalgaard and Cliver (2007a) who revised their estimate of the difference further downward to 3 nT. These authors also showed that most of the difference arose in a 6-year interval around 1957, which is the time of the move of the northern hemisphere *aa* station from Abinger to Hartland. Independently, Lockwood *et al.* (2006b) carried out tests of *aa* using the range *Ap* index which has been constructed since 1936 from 11 – 13 northern hemisphere stations, and range k indices from a number of other stations (thereby ensuring that they were comparing like-with-like). They also found a step-like change around 1957 and estimated it to be about 2 nT in magnitude. Because 1957 was only 11 years before the end of the data series

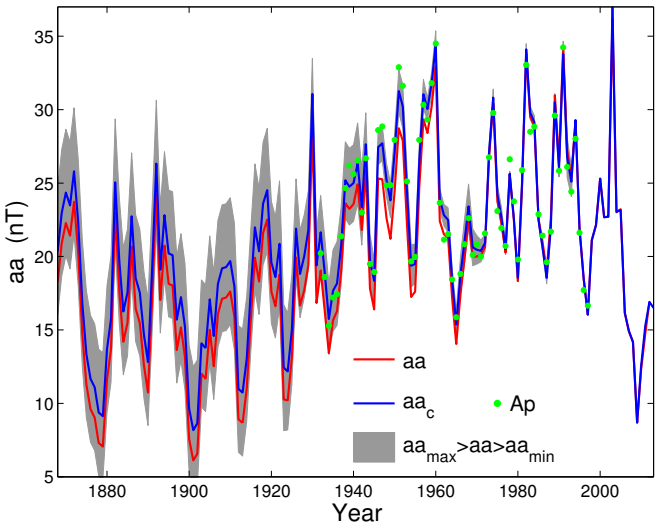

Figure 5: Comparison of annual means of the standard *aa* index (in red) and the modified version aa_C derived by Lockwood *et al.* (2006b) (in blue). The estimated maximum uncertainty, relative to modern values, is shown by the grey band. The green dots are annual means of the *Ap* index.

available to Mayaud and because in that time solar cycle 20 was rather unusual, this discontinuity in *aa* was not as apparent in the original *aa* data as it is now. Other studies also indicate that *aa* needs adjusting by about 2 nT at this date (Jarvis, 2004; Martini *et al.*, 2012). The 2 nT discontinuity estimate corresponds to an error in the drift in *aa* between cycle 14 and the space age of about 25%. Lockwood *et al.* (2006b) implemented revised calibrations between stations (the largest change needed being for 1957) and, hence, derived a revised index series aa_C. Figure 5 shows that the difference between annual means of *aa* and aa_C is generally less than 2 nT, which is considerably smaller than the range of the long-term drift in *aa* annual means over the last 150 years (approximately 12 nT at sunspot minimum and 16 nT at sunspot maximum).

Many historic datasets exist in the form of hourly mean data (or in the case of some of the earliest data, spot values within the hour) and these have recently also been used to generate indices. Until recently many were in the form of paper records in observatory yearbooks. However, in recent years many have been digitised making a valuable new extra resource for reconstruction work.

Figure 6 shoes the variation of the "median index", *m* (Lockwood *et al.*, 2006b). The construction of this index recognises that the response to global geomagnetic activity at a given observatory depends upon its magnetic local time (MLT) and, hence, on the Universal Time (UT). However, the station gives information at all UT and so rather than discard data from all but one MLT, the *m* index treats each station-UT as a separate data series. To avoid outliers having a disproportionate effect, *m* is defined as the median of all the normalised annual values for the different station-UT combinations. The black line in the upper panel of Figure 6 is *m*, which also shows a similar long-term variation to the annual means of *aa* and $IDV(1d)$ shown in Figure 4.

Figure 7 shows the *IDV* index compiled by Svalgaard and Cliver (2010). These authors take the series back to 1835, just 3 years after the establishment of the first magnetic observatory in Göttingen. This is done using a linear correlation between *IDV* proper and the *u* index. However, it must be remembered that *u* is not an inter-diurnal variability index before 1872 and Bartels did not regard all the data before this date as reliable. Figure 7 also shows the variation in *N*, the number of stations used to compile *IDV*.

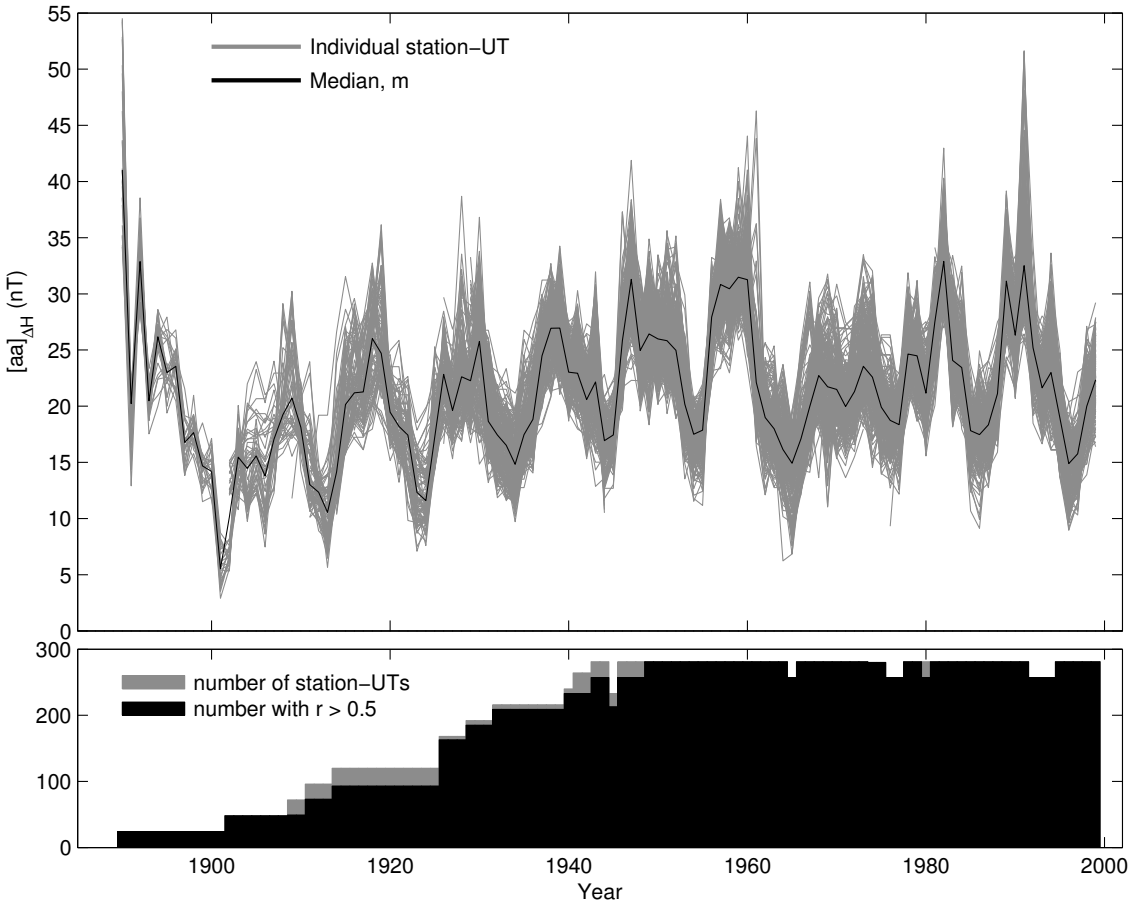

Figure 6: The median index, m. For each station at a given UT, the standard deviation of the hourly means of the horizontal component of the geomagnetic field is computed over a full year, $\sigma^H_{1\,\mathrm{yr}}$. These are then correlated with, and linearly regressed against, the annual means of aa shown in Figure 4 to yield $\sigma' = s \times \sigma^H_{1\,\mathrm{yr}} + c$. These normalisations are needed because both the sensitivity s and offset c for a station have been shown to depend on its location and on the UT hour (which, for example, alters the location of the station relative to the midnight-sector auroral oval) (Finch, 2008). Each station-UT is treated as an independent data series. The grey lines show the variations of the σ' values for all station-UTs for which the correlation coefficient exceeds 0.5 and is significant at the 2σ level. The number of station-UTs meeting this criterion is shown as a function of time by the black histogram in the lower panel. The black line in the upper panel is the median of all the available data for each year and is called the "median index", m. Image reproduced from Lockwood *et al.* (2006b).

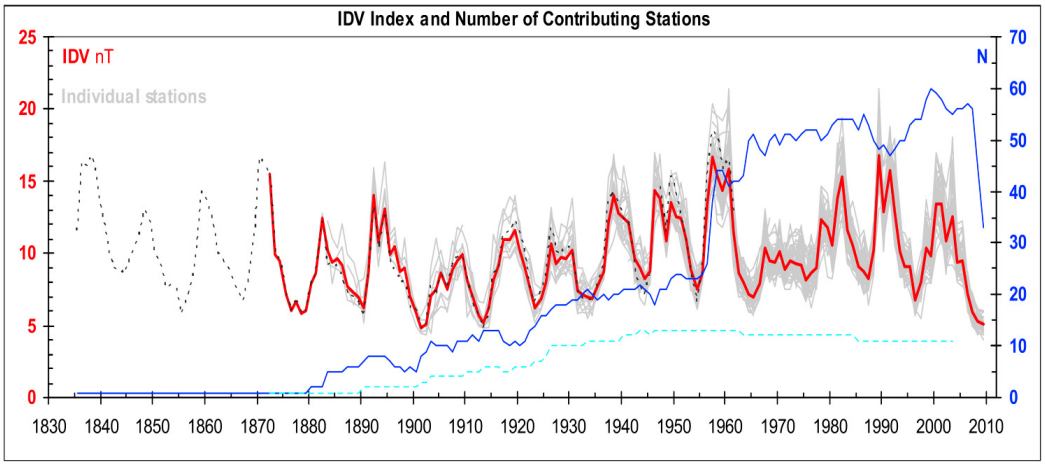

Figure 7: Annual means of the IDV index compiled by Svalgaard and Cliver (2010). The grey curves are the variations for individual stations. The red curve is the index, defined as the arithmetic mean of the median and average values of the individual station values. A few station values were very large outliers of the distribution at any one time and those that were more than five standard deviations from the average were omitted in calculating the IDV value for that year. The number of contributing stations, N, is shown by the thin blue curve. The dashed blue line is the corresponding number of stations used by Svalgaard and Cliver (2005). Bartels' u index is considered a single station and gives the dotted line extension to before 1871 using a linear regression of 1871 – 1930 data with the IDV index proper. Image reproduced by permission from Svalgaard and Cliver (2010), copyright by AGU.

Note that for both IDV and m, the number of stations used decreases as one goes back in time, which contrasts with Mayaud's philosophy for aa which was to derive a homogeneously-constructed data series. Given the potential for site-dependent errors and drifts, these indices therefore become increasingly unreliable as one goes further back in time. Svalgaard and Cliver (2010) state that "only a few (good) stations are needed for a robust determination of IDV". This is indeed a valid statement: for example, $IDV(1d)$ (shown in the bottom panel Figure 4) is based on just one station at any one time, yet for 1880 – 2013 it gives a correlation coefficient of 0.96 with IDV (Lockwood *et al.*, 2013a). However, before 1880 the correlation is considerably lower. Svalgaard and Cliver (2010) note that they had to discard some data because they were more than 5σ from the mean. This poses a dilemma if there are too few stations to define the distribution: in such cases these outliers could not be identified and one would have used them, not knowing they were in error. In other words, without sufficient other stations to compare with, one is not able to say which the "good" stations are. It therefore is inevitable that the inhomogeneous data series such as m and IDV are less reliable further back in history. Potential causes of additional uncertainty in early data are: (1) there were fewer stations; (2) measurement techniques and equipment improved with time; (3) the realisation of that urban environments were generating magnetic noise problems forced moves to quieter observing sites; and (4) earlier data tend to be spot values rather than hourly means. On the last point, Svalgaard and Cliver (2010) could find no discontinuities in the IDV data series from individual stations (unlike IHV values) when they changed from supplying spot values to hourly means. Nevertheless, it is self-evidently true that hourly means are preferable to spot values, particularly if a site is suffering from any intermittent noise problems and/or if the instrument stability is poorer.

Figures 4, 6, and 7 all show similar long-term variations, despite the fact that the indices presented differ in almost every facet of their compilation. There are, however, important differences that are discussed in Section 6. These data from geomagnetic observatories give an invaluable resource for studying solar-terrestrial physics and solar variability in the 181 years since Gauss' first observatory was established in Göttingen. In particular, we can study the variations in the solar corona and interplanetary medium that accompany the long-term sunspot variations identified by Gleissberg (1944). Feynman and Crooker (1978) studied the implications of the drift in the aa index and concluded that either the solar wind speed or the IMF had changed over the past century. The first paper to separate these two influences (using the recurrence index of Sargent, 1986, to quantify solar wind speed), thereby showing that the main change was in the magnetic field, was by Lockwood *et al.* (1999a). These authors used aa to reconstruct the unsigned open solar flux, which is the total magnetic flux leaving the top of the solar corona and entering the heliosphere. Other solar terrestrial phenomena, such as lower latitude auroras, were found to reveal the same long term changes as aa and the derived open solar flux (for example, Pulkkinen *et al.*, 2001).

4 A Note on the Importance of Understanding the Provenance of Geomagnetic Data

Before continuing, there is an important point that needs to be made about correcting and homogenising historic data. There is the potential to do much more harm than good, if corrections that are based on inadequate understanding (or, worse still, postulated theories) are allowed to modify a dataset but clear metadata and the means to reverse the changes, at any stage in the future, are not retained and made readily available. The full provenance of any one dataset is easily lost and without it such a change could be a massively retrograde step. I therefore strongly recommend that historic datasets that are re-processed should be re-named so they can be recognised for what they are and the original dataset must be retained. Hence, although at the present time it is reasonable, for example, to regard aa_C as a corrected form of aa, should something in the revised inter-calibrations in future prove to be invalid or inadequate, then scientists can readily return to the original aa data. For this reason, Lockwood *et al.* (2006b) treated aa_C as a different index to aa and gave it a new name.

A good example of the sort of problems that can arise is provided by the hourly mean H data from the Eskdalemuir station. This observatory has operated continuously since 1911, when it was established by Kew observatory on a rural and exceptionally clean magnetic site when the Kew site was rendered too noisy by the introduction of trams into west London (Harrison, 2004). There was a discontinuity at 1932 in the commonly-used set of hourly mean data from this station, which had remained un-noticed until 2004, when Mursula *et al.* (2004) and Clilverd *et al.* (2005) analyzed the inter-hour variability of Eskdalemuir data and found very small values in the early part of the 20th century. Detective work by Leif Svalgaard established that prior to 1932 the data stored in the Word Data Centre (WDC) system were 2-hour running means of the data recorded in the observatory yearbook. Such smoothing greatly influences inter-hour indices. MacMillan and Clarke (2011) have confirmed that this was indeed the case and digitised the data from the yearbook, so that all data from Eskdalemuir now available from WDC-C1 are hourly means with no running mean smoothing applied. (Users should check which dataset they are using because one problem with data that has been corrupted or massaged is that it is very hard to expunge from all datasets and bad data tends to resurface). It is not known how, when, where, or why this post-processing was carried out because the available metadata did not tell us the full provenance of the data. Presumably somebody, somewhere had believed that the noise suppression obtained by implementing a running mean was a good thing. If one used daily means of the (supposed) hourly data there would have be a some effect (as an hour of data from both the day before and the day after would be averaged in with half weight), but it would be small and the effect would be negligible on annual means. It is fair to assume that whoever implemented the smoothing never envisaged the use of the data to generate an inter-hour variability index. This example illustrates very graphically the great importance of knowing, as far as is possible, the true provenance of historic data and of all the corrections and changes that may have subsequently been applied to them. Lockwood *et al.* (2013a) have revealed a similar issue with data from Ekaterinburg by implementing an inter-correlation of hourly mean H data from a given station at different UTs as a check of data consistency: they found very high correlations around 1900, revealing that interpolation to hourly values from more sparse data had taken place.

This is a vitally important concern for reconstruction work: being overly ready to accept an adjustment is highly irresponsible as it could deny future generations of scientists the opportunity to properly exploit the data or, in a worst case scenario, seriously mislead them (Council of AGU, 2009; Vogel, 1998).

5 Solar Wind Coupling Functions, the Importance of Averaging and Allowance for Data Gaps

As discussed in Section 2.3, geomagnetic activity is enhanced when the northward component of the interplanetary magnetic field (IMF), in the GSM frame of reference, B_Z, is increasingly negative. As a result, a half-wave rectified form of B_Z is often used to predict geomagnetic activity, such as B_s, where $B_s = -B_Z$ when $B_Z \leq 0$ and $B_s = 0$ when $B_Z > 0$. Because B_s is discontinuous in slope around $B_s = 0$, a form such as $B \sin^4(\theta/2)$ is often preferred, where the IMF "clock angle" $\theta = \tan^{-1}(B_Y/B_Z)$, B_Y being the dawn-to-dusk component of the IMF in GSM. This has a very similar form to B_s at large $|B_s|$ but is continuous in slope around zero. The power density in the solar wind at Earth is dominated by the kinetic energy of the bulk flow of the particles and so the square of the solar wind velocity, $V_{\rm SW}$, is another important factor. Hence a simple "coupling function", designed to quantify the effect of the solar wind on geomagnetic activity, is $B_s \times V_{\rm SW}^2$. A great many such coupling functions have been proposed and tested. One widely-used example is the "epsilon parameter", $\epsilon = (4\pi/\mu_o) V_{\rm SW} B^2 \sin^4(\theta/2) l_o^2$ where μ_o is the magnetic permeability of free space and l_o is a scaling factor that allows for the cross-sectional area of the geomagnetic field presented to the solar wind. In practice this area reduces with increased solar wind dynamic pressure ($P_{\rm SW} = m_{\rm SW} N_{\rm SW} V_{\rm SW}^2$, where $m_{\rm SW}$ is the mean solar wind ion mass and $N_{\rm SW}$ is the number density of solar wind ions) but a constant value of $l_o = 7\,R_E$, where R_E is a mean Earth radius, is often used. However ϵ is based on the energy density in the solar wind hitting the Earth's space environment being in the form of Poynting flux, which is not correct because by far the largest energy density in the undisturbed solar wind is in the form of the ions' bulk-flow kinetic energy (which is converted into Poynting flux by the currents that flow in Earth's bow shock and magnetopause, see Cowley, 1991; Lockwood, 2004). A correct version was provided by Vasyluinas *et al.* (1982) who applied dimensional analysis as well as the energy flow equations and used pressure balance on a hemispherical dayside magnetopause to compute l_o. From this they derived the coupling function P_α, which is the power coupled into the magnetosphere. It is the product of the power density in the solar wind, times the cross sectional area of the magnetosphere presented to the solar wind, times the fraction of the incident power that crosses the magnetopause, t_r

$$P_\alpha = (m_{\rm SW} N_{\rm SW} V_{\rm SW}^3)/2 \times (\pi l_o^2) \times t_r \,. \tag{1}$$

From pressure balance at the nose of the magnetosphere, and assuming the dayside magnetosphere is hemispherical in shape we have

$$l_o = k_1 \left[M_E^2/(P_{\rm SW} \mu_o) \right]^{1/6} , \tag{2}$$

where M_E is the Earth's magnetic moment and k_1 is the blunt-nose shape factor for flow around the magnetosphere. Vasyluinas *et al.* (1982) noted that the transfer function t_r must be dimensionless and proposed a form

$$t_r = k_2 M_A^{-2\alpha} \sin^4(\theta/2) \,, \tag{3}$$

where α is a free fit parameter which arises from the unknown dependence of the coupling on the solar wind Alfvén Mach number, $M_{\rm A}$. Combining Equations (1), (2), and (3) yields

$$P_\alpha = \left[k_1 k_2 \pi/(2\mu_o^{(1/3-\alpha)}) \right] M_E^{2/3} m_{\rm SW}^{(2/3-\alpha)} B^{2\alpha} N_{\rm SW}^{(2/3-\alpha)} V_{\rm SW}^{(7/3-\alpha)} \sin^4(\theta/2) \,. \tag{4}$$

Finch and Lockwood (2007) studied the interplanetary medium coupling functions P_α, $V_{\rm SW}^2 B$, $V_{\rm SW}^2 B_S$, ϵ, $V_{\rm SW}^2$, $P_{\rm SW}$, and B by evaluating their correlations with various geomagnetic activity indices on a range on averaging timescales, T between 1 day and 1 year. The results for the Am global index are shown in Figure 8.

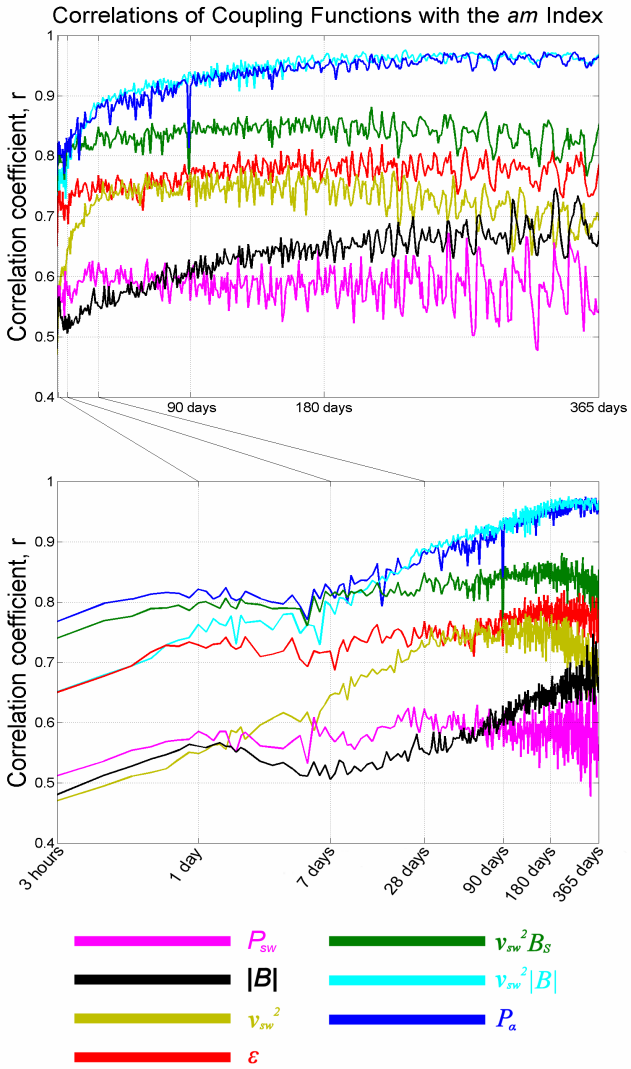

Figure 8: Correlations between the *Am* geomagnetic activity index and a number of solar wind coupling functions, as a function of averaging timescale, T. Upper and lower graphs are identical, other than that the upper graph displays timescale linearly, whereas the lower graph displays it logarithmically. The coloured lines give the results for: P_α (dark blue), $V_{\rm SW}^2 B$ (light blue), $V_{\rm SW}^2 B_S$ (green), ϵ (red), $V_{\rm SW}^2$ (olive), $P_{\rm SW}$ (magenta), and B (black). Image reproduced by permission from Finch and Lockwood (2007), copyright by EGU.

A factor that should not be neglected is that although the geomagnetic data are essentially continuous, the same is far from true of the interplanetary data. Since 1995 the WIND and the ACE spacecraft have provided almost 100% coverage, but before then coverage had sometimes been lower than 50% in any one year. Finch and Lockwood (2007) showed that ignoring these data gaps has a considerable effect at given averaging timescale T and can even change which coupling function performs best. Hence before making the correlation, Finch and Lockwood (2007) piece-wise removed both interplanetary and geomagnetic data for which there was a gap in the interplanetary data of duration one hour or greater during a 3-hour geomagnetic data interval (allowing for the predicted propagation lag between the interplanetary monitoring spacecraft and the dayside magnetopause). Figure 8 shows that for the full range of T, P_α (as given by Equation (4) and shown in dark blue) performs best for the range index Am, although for $T > 27$ days the much simpler function $V_{SW}^2 B$ (light blue) gives correlations as high (or even slightly higher). The ϵ parameter (red) performs as well as $V_{SW}^2 B$ (but less well than P_α and $V_{SW}^2 B_S$) at low T and is considerably poorer at high T, reflecting its nonphysical basis. For the case of Am shown, functions that do not combine both IMF and solar wind speed (V_{SW}^2 in olive, P_{SW} in magenta, and B in black) do not perform as well as those that do. There has been much discussion about the precise form of coupling function that performs best, but these discussions almost invariably neglect the facts that this conclusion depends on T and on which activity index is considered. The importance of this is discussed in Section 6. Note that in Figure 8 the high-performing indices P_α and $V_{SW}^2 B$ reach correlation coefficients near 0.97 at $T = 1$ yr and that these do not fluctuate with the precise T value used to anything like the same extent as do the others.

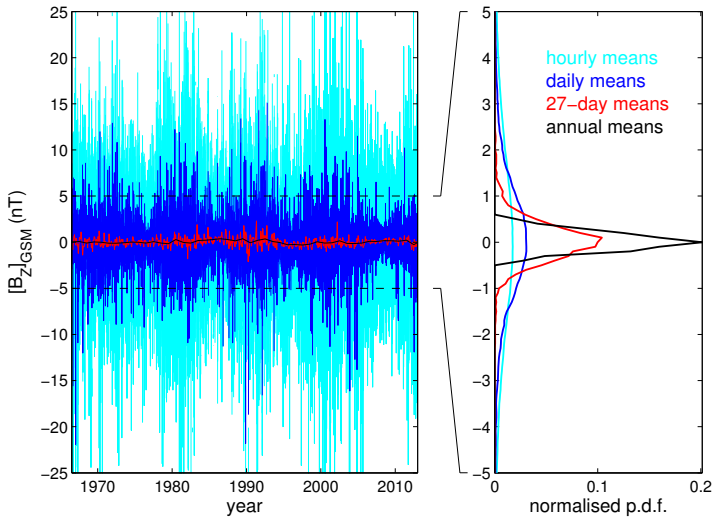

Figure 9: The northward component of the near-Earth interplanetary magnetic field (IMF) in the GSM frame, B_Z, from the OMNI2 composite dataset. The left-hand plot shows the temporal variations of $\langle B_Z \rangle_T$ for 1966 – 2013 at different averaging timescales, T. The right-hand plot shows the corresponding probability density functions over the same interval. Note the difference in the vertical scales. In both panels light blue is for $T = 1$ h, dark blue for $T = 1$ day, red for $T = 27$ days and black for $T = 1$ yr.

One noticeable feature of Figure 8 is that for large T, $V_{SW}^2 B$ performs as well as P_α which is interesting because it does not contain the $\sin^4(\theta/2)$ IMF orientation factor that equation 4 shows is part of P_α. Figures 9 and 10 explain why this is the case. The left-hand panel of Figure 9 shows the northward component of the IMF in the GSM reference frame, B_Z between 1966 and 2012 (inclusive). The different colours identify the timescale T on which the data are averaged before

they are plotted. For $T = 1$ h (light blue), values vary between –30 nT and +30 nT (off scale), and periods of smaller and larger excursions, both positive and negative, are seen (corresponding to the peaks and minima of the solar cycle). The same is true for $T = 1$ day (dark blue) but the range of variation is reduced as intervals of opposite B_Z cancel to a great extent for the larger T. For $T = 27$ days (in red) the fluctuation level is very small and it has almost disappeared for $T = 1$ yr (black). The distributions of B_Z values over the interval are shown for each case in the right-hand plot. It can be seen that the averaging almost completely removes the orientation factor such that B_Z tends to zero for large T.

The effect of averaging is demonstrated by Figure 10 which shows scatter plots of $B\sin^4(\theta/2)$ as a function of the IMF magnitude B for the 1966 – 2012 data. Part (a) is for hourly observations. It can be seen that there is large scatter between $B\sin^4(\theta/2) = B$ (when the field points directly southward so $\sin^4(\theta/2) = 1$) and $B\sin^4(\theta/2) = 0$ (when the field points directly northward so $\sin^4(\theta/2) = 0$). Part (b) is the same for 1-year averages. In this case there is a good linear relationship, with some scatter. Hence, on timescales of $T = 1$ yr, the IMF orientation factor is averaged out and the average southward IMF component and, hence, the level of geomagnetic activity, is proportional to B, as first noted by Stamper *et al.* (1999). This is the basic reason that we are able to make deductions about B from geomagnetic activity when averaging is done on annual timescales. There is some information that could be extracted at higher time resolution, but Figure 9 shows that even for $T = 27$ days scatter will be introduced because this T is not sufficient to average out as much of the IMF orientation factor and at yet smaller T this scatter would render the results completely meaningless. In this review we restrict our attention to using $T = 1$ yr, for which the orientation factor is almost completely averaged out and for which correlations between the better coupling functions and geomagnetic activity of 0.97 can be obtained (as shown by Figure 8). Part (c) of Figure 10 shows the distribution of annual values of the ratio $B/[B\sin^4(\theta/2)]$. The mean value of this distribution is 3.251 and the standard deviation is 0.369. This distribution will be used in Section 9.4 in a quantitative analysis of the uncertainties in reconstructions.

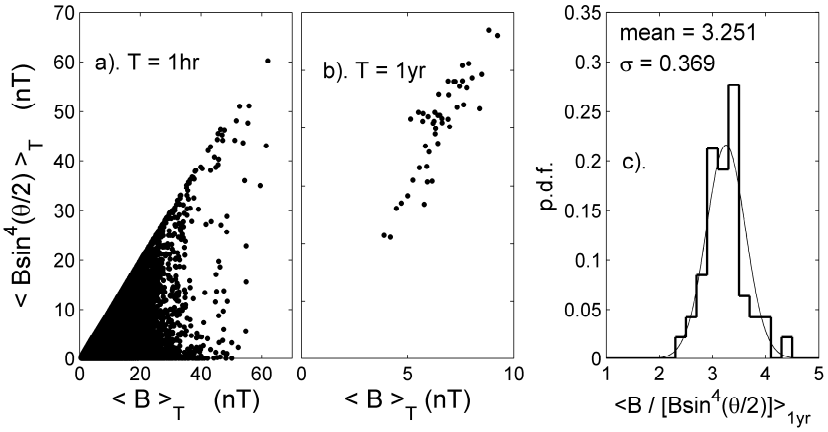

Figure 10: Scatter plots of the half-wave rectified southward component of the IMF in the GSM frame, $B\sin^4(\theta/2)$ as a function of the IMF magnitude, B, for (a) hourly observations and (b) 1-year averages. Panel (c) shows a histogram of the distribution of annual values of the ratio $B/[B\sin^4(\theta/2)]$, which has a mean value of 3.251 and a standard deviation of 0.369. The thin line is the normal distribution with the same mean and standard deviation.

In addition, in order to derive the open solar flux, the modulus of the radial component of the IMF, away from the Sun (B_r, which is the same as $-B_X$ in the GSM frame) is usually used (see Section 7.3). $|B_r|$ can be obtained from B, again because of the effect of averaging. Because in

one year roughly as much "Toward" IMF ($B_r < 0$) flux will be seen as "Away" ($B_r > 0$) flux, $\langle B_r \rangle_T$ will tend to zero when averaged over $T = 1$ yr and so $|B_r|$, or some equivalent which does not cancel Toward and Away flux (see Section 7.5), is needed. The orientation angles of the IMF, both the clock angle θ and the garden-hose angle $\zeta = \tan^{-1}(B_X/B_Y)$, vary considerably on short time scales. Figure 9 shows that the long-term average of θ is 90° and that of ζ is given by Parker spiral theory (Parker, 1958, 1963) which predicts for near-Earth interplanetary space

$$\zeta = \tan^{-1}(V_{\rm SW}/r_1\omega) = \sin^{-1}(|B_r|/B)\,, \tag{5}$$

where r_1 is the mean Earth-Sun distance ($r_1 = 1$ Astronomical Unit, AU), and ω is the angular rotation velocity of the solar atmosphere with respect to the fixed stars. Equation (5) shows that the ratio $|B_r|/B$ can be predicted for a given $V_{\rm SW}$. The left hand panel of Figure 11 shows a scatter plot of the values of $|B_r|/B$ predicted by Equation (5) against the observed values for hourly means ($T = 1$ h). It can be seen that the large variations of θ and ζ on hourly timescales mean that there is no relationship between the observed and predicted values. On the other hand, the right hand plot shows the scatter plot for annual means. (Note that the modulus of B_r has here been taken of means over 1 day, i.e., $\langle |B_r|_{1\,\rm d} \rangle_{1\,\rm yr}$ is used). For this timescale there is a linear relationship between the observed and predicted values. The observed values are lower than the predicted ones because of the use of $T = 1$ day in taking the modulus which means there is some cancellation of Toward and Away field: this issue is discussed further in Section 7.5. Figure 11 demonstrates that Parker spiral theory can be used to predict $|B_r|$ from B if the solar wind speed $V_{\rm SW}$ is known and an annual averaging timescale is used.

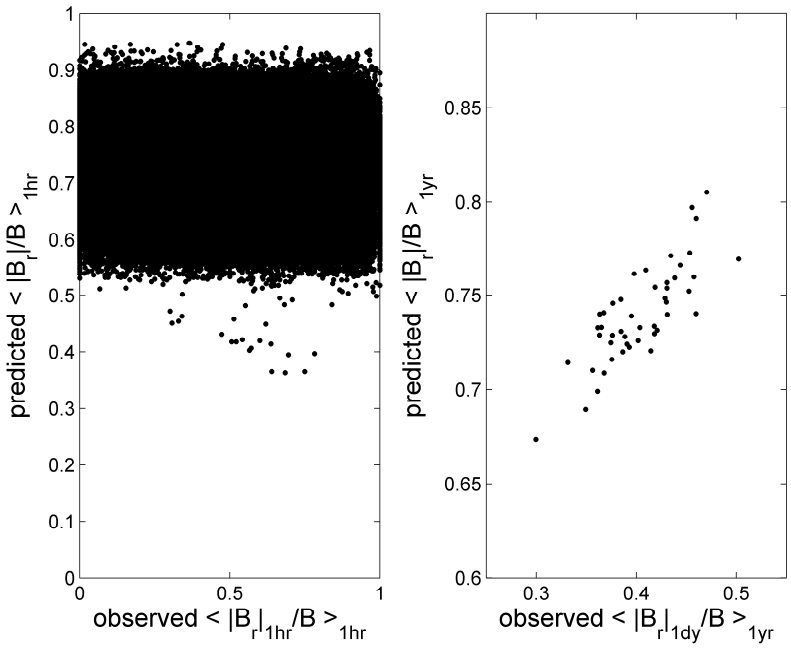

Figure 11: Scatter plots of the predicted ratio $|B_r|/B$ from Parker spiral theory, where B is the IMF magnitude and $|B_r|$ is the modulus of its radial component, as a function of the observed value of that ratio. The left-hand plot is for hourly observations, the right-hand plot for 1-year averages. In the case of the annual means, the modulus of daily means of B_r observations, i.e., $\langle |B_r|_{1\,\rm d} \rangle_{1\,\rm yr}$, is used.

6 Differences Between Range and Hourly Mean Geomagnetic Data and the Effect of Solar Wind Speed

Figure 3 shows that there is a consistent difference between interdiurnal variation indices and the range indices. The interdiurnal variation indices, such as IDV, $IDV(1d)$, and m all correlate best with the IMF field strength B on annual averaging timescales whereas mid-latitude range indices (aa, Ap, Kp, Am – and its northern and southern hemisphere components An and As) all correlate best with a coupling function close to BV^2.

We should not expect these two classes of indices to behave in the same way. Consider a "steady convection event" (see, for example, Lockwood *et al.*, 2009a, and references therein) lasting, for example, 24 hours in which DP2 is enhanced but DP1 is not because of the lack of substorms. There has been debate about whether or not the ring current is enhanced during such events (Pulkkinen, 2007), the outcome of which appears to be that although ring current enhancements are weaker because of the lack of substorms, they are still present (for example Zhou *et al.*, 2003). Given that interdiurnal variability indices appear to be particularly sensitive to the ring current and/or the DP2 system, steady convection events will influence these indices much more than range indices (which respond most strongly to the DP1 currents). On the other hand, because the substorm cycle of energy storage in the magnetospheric tail (associated with DP2) and its explosive release (associated with DP1) generally takes place within 3-hour intervals, we would expect to see strong signatures of substorms in the range indices, but weaker ones in interdiurnal range indices. This line of argument suggests that the ratio of sensitivities to DP1 and DP2/Dst may be greater for range indices than for interdiurnal variability indices. In this section, we will discuss evidence that this is the case and show that it causes them to have different responses to variations in the interplanetary medium, such that the optimum coupling functions are not the same.

The paper by Finch *et al.* (2008) provides a very important insight. These authors devised the "sigma-H" σ_n^H indices for each station, based on hourly mean data (see Section 2.2.6). The series of $\langle\sigma_{28}^H\rangle_{1\,\mathrm{yr}}$ values for each station were correlated with three interplanetary parameters: B, V_{SW}, and $V_{\mathrm{SW}}^2 B$. The zero-lag correlation coefficients are plotted in Figure 12 as a function of the modulus of the station's invariant geomagnetic latitude, $|\Lambda|$. The top panel of Figure 12 shows that the correlation coefficient for B is very high (~ 0.9, significant at the 2σ level) except at auroral latitudes (between the two vertical dashed lines) where it falls to a minimum of about 0.6. On the other hand, the correlation with V_{SW} (middle panel) is low outside the auroral oval (generally below about 0.4), but rises to a peak of about 0.85 within the oval. The bottom panel shows that $V_{\mathrm{SW}}^2 B$ is high and ≈ 0.8 outside the oval and ≈ 0.9 within it. Thus the influence of V_{SW} arises in the auroral oval, making $V_{\mathrm{SW}}^2 B$ correlate best there but elsewhere B alone provides the highest correlation.

Figure 13 shows the MLT at which the peak correlations within the auroral oval occur. The primary source of the correlation with V_{SW} occurs in the midnight MLT sector. This is when and where westward electrojet of the DP1 system is most likely to be detected (Tomita *et al.*, 2011). The maximum correlation with B can occur at any time except in the midnight sector. Therefore, it is clear that the correlation with the solar wind velocity at yearly time scales is linked to the storage-release system of the magnetotail and the westward auroral electrojet of the DP1 current system. At all other locations the correlation is better with B and appears to be more associated with the directly-driven DP2 currents and/or with the ring current.

This finding makes good sense, physically. The westward auroral electrojet of the DP1 current system is part of the "substorm current wedge" (see Figure 14), in which the dawn-to-dusk current in the near-Earth edge of the magnetospheric cross-tail current sheet is diverted during substorm expansion through the midnight-sector auroral oval (McPherron *et al.*, 1973). In this schematic, (a) and (b) are for a substorm growth phase, (c) and (d) are for a substorm expansion phase. The undisturbed IMF, B is draped over the nose of the magnetosphere by the slowing effect of the

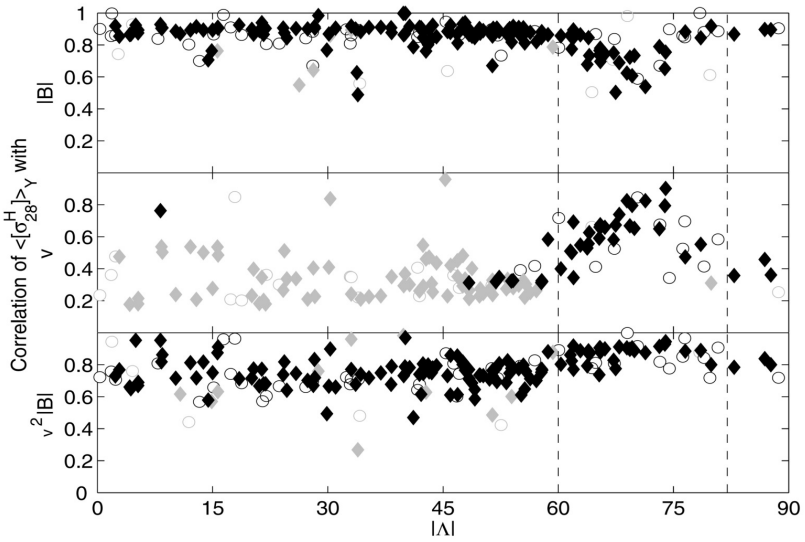

Figure 12: Correlation coefficients of the annual means of the coupling functions (top) IMF magnitude B, (middle) $V_{\rm SW}$, and (bottom) $V_{\rm SW}^2 B$ with σ_{28}^H indices from a variety of magnetic observatories, shown as a function of the modulus of their invariant magnetic latitude $|\Lambda|$. Solid diamonds are for northern hemisphere stations and open circles for southern. Black and grey symbols are for correlations that are significant at the 2σ and 1σ levels, respectively. The vertical dashed lines show the auroral oval between absolute invariant latitudes of 60° and 82°. Image reproduced by permission from Finch *et al.* (2008), copyright by AGU.

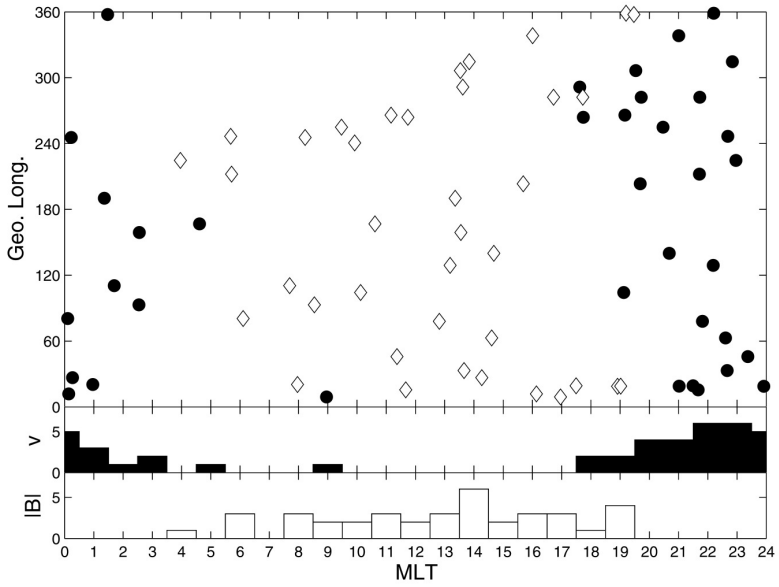

Figure 13: MLT dependence of maximum correlations of $\langle\sigma_{28}^H\rangle_{1\,\rm yr}$ for stations shown in Figure 12 with $60° < |\Lambda| < 82°$. In the top panel, filled circles and open diamonds show the MLT (magnetic local time) – Geographic longitude coordinates of the maximum correlation of $\langle\sigma_{28}^H\rangle_{1\,\rm yr}$ with $|V_{\rm SW}|$ and B, respectively. The bottom two panels show the histograms of the MLT of the maxima in the correlations for $|V_{\rm SW}|$ (in black) and B (in white). Image reproduced by permission from Finch *et al.* (2008), copyright by AGU.

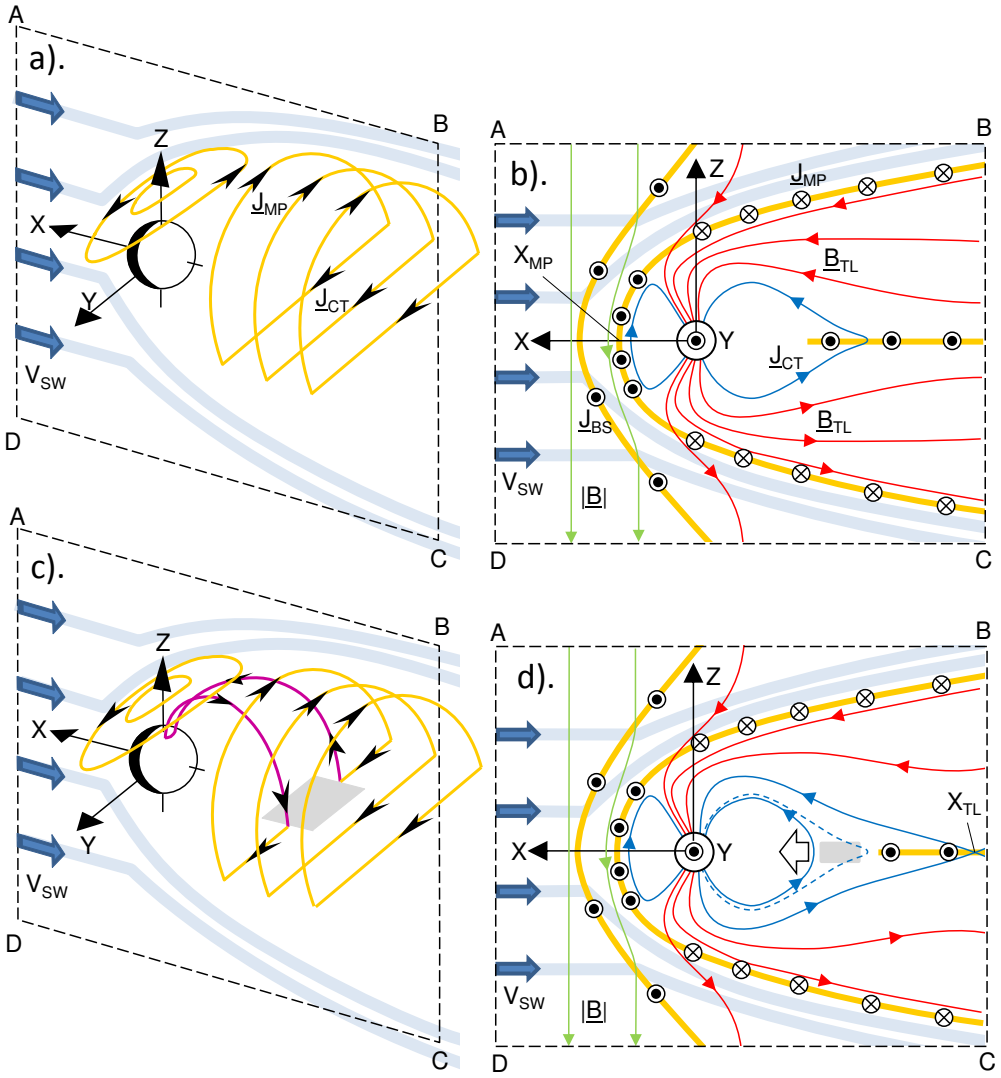

Figure 14: Schematic of Earth's magnetospheric current systems. Parts (b) and (d) are views of the noon-midnight cross-section of the magnetosphere, ABDC, from the ecliptic on dusk side of the Earth and show current sheets in orange, interplanetary field lines in green, open geomagnetic field lines (that thread the magnetopause) in red and closed field lines (that do not) in blue. Parts (a) and (c) are views of the northern-hemisphere currents in the tail and dayside magnetopause from high northern latitudes in the pre-midnight sector. In all panels X, Y, and Z are the axes of the GSM reference frame. Parts (a) and (b) are for a substorm growth phase, whereas (c) and (d) are for a substorm expansion phase. B is the undisturbed IMF and J_{BS} is the current in the Bow Shock. In all panels magnetic reconnection is taking place at X_{MP} in the dayside magnetopause. The tail lobe field is B_{TL} and the current in the cross-tail current sheet (J_{CT}) is disrupted in the grey areas in (c) and (d). The mauve currents in (c) are the "substorm current wedge" within which the magnetospheric field lines "dipolarise" at onset (the dashed line in (d) shows the stretched field line before onset and the arrow the associated sunward convection surge of the frozen-in plasma). The "near-Earth neutral line", X_{TL}, is also shown in (d).

bow shock (currents J_{BS}). The IMF in this schematic points due south in the GSM frame which favours magnetic reconnection at X_{MP}, driven by the large magnetic shear across the dayside magnetopause current sheet. This generates open magnetospheric flux that is appended to the tail by the solar wind flow, causing rises in the tail lobe field B_{TL} and in the current in the near-Earth cross-tail current sheet (J_{CT}) during the growth phase. This rise comes to an end at the onset of the expansion phase, when the Earthward edge of the cross-tail current is disrupted in the area shown in grey in (c) and (d). This current is diverted down post-midnight field lines, along the westward electrojet in the ionosphere and back up pre-midnight field lines. This path is shown in mauve in (c) and is called the "substorm current wedge" which gives the magnetic disturbances classed as DP1. Within the wedge the field dipolarises as shown in (d) (the dashed line being the stretched field line before onset and the white arrow the sunward convection surge of the frozen-in plasma). Part (d) also shows open flux now being destroyed by reconnection at a "near-Earth neutral line", X_{TL}. The DP2 currents that dominate during the growth phase are driven by the reconnection in the dayside magnetopause and so depend strongly on the southward IMF. In comparison, the cross-tail current diverted into the auroral electrojet (to give the DP1 disturbances) is set by B_{TL} which depends on both the total open magnetospheric flux, (and, hence, also on the IMF) and the solar wind dynamic pressure ($\propto V_{SW}^2$) which squeezes the tail where the current disruption takes place. This is because the cross-tail current disruption occurs close to the Earth where the magnetotail is still flaring (i.e., its radius is increasing with distance away from Earth) which means that enhanced dynamic pressure P_{SW} caused by enhanced solar wind velocity V_{SW} can squeeze the tail at this location and so increase the field in both tail lobes, giving a higher magnetic shear across the cross-tail current sheet for a given amount of open magnetospheric flux in the tail lobes. Indeed, that there must be this V_{SW}^2 dependence in substorm-related phenomena can be seen by considering what happens further down the tail, at greater negative X coordinates. Here the tail reaches its maximum, asymptotic radius so is no longer flaring, so the magnetopause becomes parallel to the solar wind flow and P_{SW} has no influence. The magnetopause location is here set by pressure balance between the magnetic pressure that dominates in the tail lobes, $B_{TL}^2/(2\mu_o)$, and the static pressure of the solar wind, dominated by the thermal pressure of the particles, ($N_{SW}kT_{SW}$, where T_{SW} is the solar wind temperature). Substorms occur because of the growth of open flux in the tail. Because in the far tail, B_{TL} is set by pressure balance with the static pressure in the interplanetary medium, adding more open flux causes the far tail radius to increase, but B_{TL} and J_{CT} remain constant. Only closer to the Earth, where the tail is flaring, does the solar wind dynamic pressure act to constrain the tail radius so that the accumulation of open magnetospheric flux there causes a rise in B_{TL} and J_{CT}. Hence, substorm phenomena such as the current disruption and the formation of the near-Earth neutral line must occur relatively close to the Earth and must have a dependence on solar wind dynamic pressure (as inferred from observations by Karlsson *et al.*, 2000).

The conclusion from the above considerations is that the current available to be diverted when the current wedge forms is higher when V_{SW} is high. The net result is that the DP1 currents have a dependence on V_{SW}^2. Recalling the conclusion, discussed earlier in this section, that range indices are likely to be influenced by the DP1 currents to a much greater extent than interdiurnal variation indices, we would expect the optimum coupling function to have a dependence on V_{SW}^2 for range indices (and for IHV) that is not seen for interdiurnal variation indices such as IDV and $IDV(1d)$.

That this is indeed the case for IHV and IDV was first noted and exploited by Svalgaard *et al.* (2003) and has subsequently been used by Svalgaard and Cliver (2007a), Rouillard *et al.* (2007), and Lockwood *et al.* (2009d). The difference for aa_C and m is demonstrated by Figure 15. Lockwood *et al.* (2009d) investigated the correlations of both aa_C and m with the general form $B^pV_{SW}^n$ and the exponents p and n giving peak correlation were quantified using the Nelder–Mead simplex search method (Nelder and Mead, 1965). Results were almost identical if the minimum

r.m.s. fit residual was searched for. In both cases, p was extremely close to unity (as used in Figure 3) and n was found to be 2 for aa_C and 0.3 for m. Thus, as predicted above, aa_C is more dependent on $V_{\rm SW}$ than m. Figure 16 underlines that the difference in the exponent n for the two cases is significant. If the exponent were to be the same in the two cases its optimum value would be about $n = 1.6$ for which the combined significance of $(1 - S_{aac}) \times (1 - S_m)$ peaks at 13%, so the difference between the exponents n in the two cases is significant at the 87% level. Plots equivalent to Figure 15 and Figure 16 reveal this difference is even more significant for the combinations IDV and IHV, IDV and aa_C, and $IDV(1d)$ and aa_C.

This difference between the dependence of the two geomagnetic indices on $V_{\rm SW}$ is extremely useful. Using the best-fit regressions such as those shown in Figure 15, the variations of both $BV_{\rm SW}^2$ and $BV_{\rm SW}^{0.3}$ can be derived and, hence, the variations of both B and $V_{\rm SW}$ can be obtained.

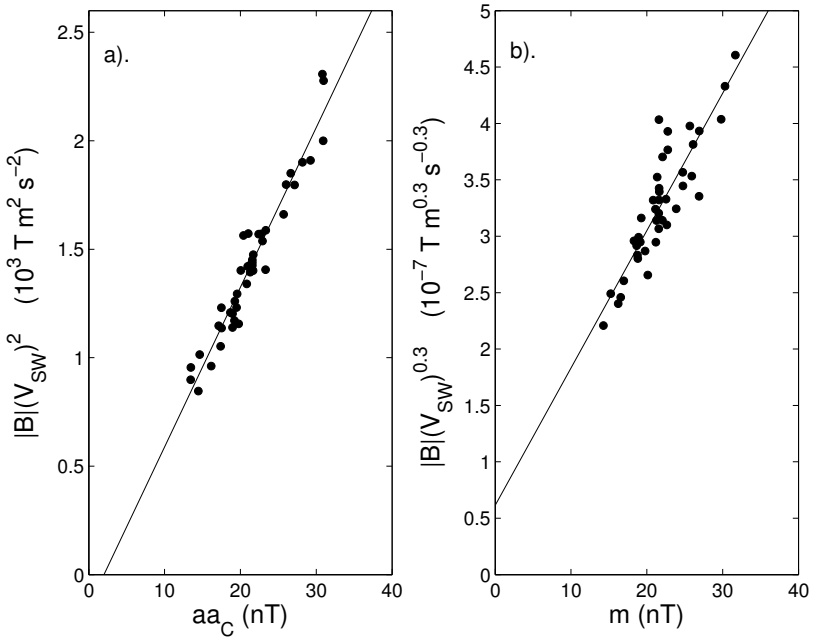

Figure 15: Scatter plots and best fit regressions of annual means of (left) aa_C against $BV_{\rm SW}^2$ and (right) m against $BV_{\rm SW}^{0.3}$. The best-fit regression lines are derived using the Bayesian least-squares regression fit procedure described by Rouillard *et al.* (2007). The correlation for aa_C with $BV_{\rm SW}^2$ is $r_{aac} = 0.97$ and using the autoregressive AR-1 red noise model this correlation is found to be significant at greater than the 10^{-5} level. The correlation of m with $BV_{\rm SW}^{0.3}$ is $r_m = 0.89$ with significance exceeding the $5{\times}10^{-5}$ level. Image reproduced by permission from Lockwood *et al.* (2009d), copyright by AAS.

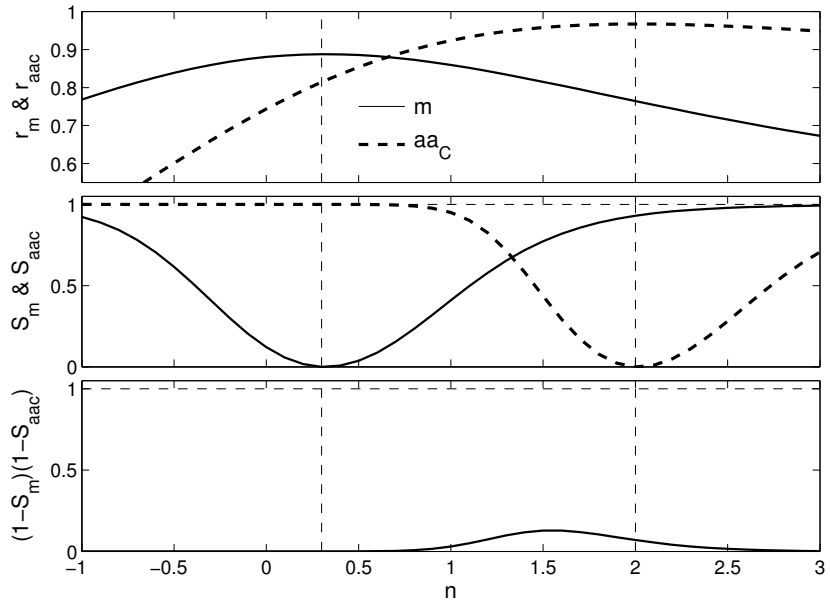

Figure 16: *Top:* Correlograms showing the correlation coefficients r_{aac} (between aa_C and $BV_{\rm SW}^n$, dashed line) and r_m (between m and $BV_{\rm SW}^n$, solid line) as a function of the exponent n. The vertical dashed lines mark the peaks at $n = 2$ and $n = 0.3$, respectively. *Middle:* the significances S_{aac} and S_m (dashed and solid lines, respectively) of the differences between the correlations at general n and the peak values, as evaluated using the Fischer-Z transform and (bottom) the joint probability $(1 - S_{aac}) \times (1 - S_m)$. Image reproduced by permission from Lockwood *et al.* (2009d), copyright by AAS.

7 Open Solar Flux, PFSS, and the Ulysses Result

7.1 Open solar flux

The open solar flux, F_S, is the magnetic flux leaving the top of the solar atmosphere and entering the heliosphere. A notional surface, the "coronal source surface" is envisaged at the top of the corona, which is everywhere perpendicular to the field, and F_S is the flux threading that surface and so is also called the "coronal source flux". This review (unless otherwise stated) considers the "signed" open flux which means that only the flux of one polarity (inward or outward) is quantified. If we assume that Maxwell's equations hold such that there are not significant numbers of magnetic monopoles inside the coronal source surface, then the inward and outward fluxes are equal and the "unsigned" open flux is simply twice the signed value. The source surface is usually taken to be a heliocentric sphere of radius $2.5\,R_\odot$, where $R_\odot$ is a mean solar radius (Arge and Pizzo, 2000).

7.2 The potential field source surface (PFSS) method

Prior to the Ulysses mission, F_S could only be evaluated from solar magnetograms using "potential field source surface" (PFSS) modelling (Schatten *et al.*, 1969; Altschuler and Newkirk, 1969; Schatten, 1999). In this method, photospheric magnetic fields observed by a magnetograph are mapped through the solar corona to the source surface with a number of assumptions. This involves solving Laplace's equation within an annular volume above the photosphere in terms of a spherical harmonic expansion, the coefficients of which are derived from Carrington maps of the photospheric magnetic field (i.e., maps assembled over an entire solar rotation from magnetograms

recorded by magnetographs on Earth's surface or on board a spacecraft in orbit around the L1 Lagrange point). Two major assumptions are that there are no temporal variations within the 27 days taken to build up the map and that there are no currents in the corona (these are neglected so as to allow unique solutions in closed form). To eliminate the possibility that such simple harmonic expansions would result in all of the magnetic field lines returning to the Sun within a small heliospheric distance, the coronal field was required to become radial at the outer boundary, the source surface. Despite its many assumptions and obvious limitations, PFSS has been very successful in the study of a wide range of solar and heliospheric phenomena, including: coronal structure as seen during eclipses (e.g., Smith and Schatten, 1970), end-to-end modelling of Earth-impacting coronal mass ejections (CMEs, e.g., Luhmann *et al.*, 2004), coronal null points and CME release (e.g., Cook *et al.*, 2009), interplanetary magnetic fields (e.g., Burlaga *et al.*, 1978), heliospheric current sheet structure (e.g., Hoeksema *et al.*, 1982), waves in the corona (e.g., Uchida *et al.*, 1973), solar wind acceleration (e.g., Neugebauer *et al.*, 1998; Marsch, 1999), stellar coronal fields (e.g., Jardine *et al.*, 2002), coronal hole and fast solar wind stream evolution (e.g., Wang and Sheeley Jr, 1990), co-rotating interaction regions and associated cosmic ray modulation (e.g., Rouillard *et al.*, 2007), solar wind speed prediction (e.g., Arge *et al.*, 2002), solar wind density structure (e.g., Rouillard *et al.*, 2010), pseudostreamers (e.g., Owens *et al.*, 2013), and quantifying the open solar flux (discussed below). The method has also generated results that compare well with images that reveal field line structure in the corona and with the results of MHD modelling (see the Living Review by Mackay and Yeates, 2012).

7.3 Ulysses observations

The Ulysses spacecraft is the first to carry out a comprehensive survey the magnetic field in heliosphere outside the ecliptic plane: its orbit covers a wide range of heliospheric distances, r and an almost full range of heliographic latitudes, Λ. This mission has generated a vitally important result in that it found that the average radial field of the heliospheric field was independent of Λ. This result was first found to apply as the satellite passed from the ecliptic plane to over the southern solar pole (Smith and Balogh, 1995; Balogh *et al.*, 1995). Subsequently, the result was confirmed by the pole-to-pole "fast latitude scans" during the perihelion passes of the spacecraft (Smith and Balogh, 1995; Smith *et al.*, 2001; Smith and Balogh, 2003; Smith *et al.*, 2003). As pointed out by Smith (2011, 2013), this "Ulysses result" was initially derived by averaging the radial field over the inferred Toward and Away sectors of the heliospheric field; however, Lockwood *et al.* (1999b), Lockwood (2004), Lockwood *et al.* (2004), and Lockwood *et al.* (2009b) have shown that the invariance with Λ was also found if the modulus of the radial field was employed over fixed averaging intervals. The differences between, and complementarity of, these two approaches are discussed in Section 7.5.

Figure 17 illustrates the Ulysses result by showing data from the third perihelion pass of Ulysses and compares them to simultaneous data from the ACE spacecraft. The top panel clearly defines the two large polar coronal holes seen by Ulysses (where V_{SW} is large), separated by a single streamer belt around perihelion (minimum r, see the black line in the bottom panel) where V_{SW} is lower. This latitudinal structure is characteristic of the sunspot minimum conditions prevailing during this pass and Ulysses passed from 80° in the south to 80° in the north (black line, panel f). Note that perihelion (minimum r_U, panel g) is within the streamer belt at $r_U \approx 1.4\,\mathrm{AU} > r_1 = 1\,\mathrm{AU}$, so when both are in the streamer belt, Ulysses is at greater r than ACE by about 0.4 AU. The panel b emphasises the variability of V_{SW} is greater within the streamer belt. Panel c shows the dominant field orientation changes from away to toward as the streamer belt is crossed and panel d shows the modulus of the radial field seen by the two craft. Note that in this plot, the influence of the timeconstant T (on which the modulus is taken) has been eliminated by using a range of T down to 1 second and fitting the results so the asymptotic value at $T \to 0$ can be calculated (see

Figure 2 of Lockwood *et al.*, 2009b): this means that there is no cancellation of toward and away field in these values. The data have been normalised to $r = r_1$ using an r-squared dependence (i.e., the Ulysses values have been multiplied by $(r_U/r_1)^2$). It can be seen from Figure 17(d) that Ulysses observed almost the same range-normalised radial field as seen simultaneously by the ACE craft: there is a weak downward trend in both, owing to the pass taking place while the solar cycle was still declining slightly.

The first and third Ulysses perihelion passes were at sunspot minimum, but the second was close to sunspot maximum. That the Ulysses result applied in all three (Lockwood *et al.*, 2004, 2009b) is an important demonstration of its generality. It is explained by the low plasma beta of the solar wind just after leaving the coronal source surface, i.e., the total magnetoplasma pressure is dominated by the magnetic pressure. This results in slightly non-radial flow close to the Sun, which smoothes out differences in the tangential pressure and, as this is dominated by $|B_r|^2/(2\mu_o)$, this renders the magnitude of the radial magnetic field, $|B_r|$ constant in latitude (Suess and Smith, 1996; Suess *et al.*, 1996, 1998). It means that the signed (of one radial field polarity) open solar flux threading a heliocentric sphere of radius r, $F(r)$ can be computed using

$$F(r) = (4\pi r^2)\langle |B_r(r)|_T \rangle_{27\,\mathrm{d}}/2\,. \tag{6}$$

Averaging over 27 days ensures that any variations with Carrington longitude are averaged out.

7.4 Excess flux

In Figure 17 there is a slight difference between the derived flux threading the sphere on which ACE sits and that threading the sphere on which Ulysses sits and this is shown by the grey area in panel (e). Lockwood *et al.* (2009b,c) have termed this difference the "excess flux", defined for general r as:

$$\Delta F = F(r) - F(r_1) = 2\pi \times \left[r_U^2 \langle |B_r(r)|_T \rangle - r_1^2 \langle |B_r(r_1)|_T \rangle_{27\,\mathrm{d}}\right] \tag{7}$$

Note that in Figure 17, the timescale on which the modulus is taken is $T \to 0$. From comparison of Figures 17(e) and 17(a), it can be seen that ΔF is larger in the streamer belt than within the polar coronal holes.

Owens *et al.* (2008a) surveyed Carrington-rotation means of magnetic field data recorded in the heliosphere (from 14 different spacecraft in different orbits) and compared them to the data taken simultaneously in near-Earth space. In their survey the modulus of B_r was taken of averages over $T = 1$ h intervals (i.e., $|B_r|_{1\,\mathrm{h}}$). The ΔF values, computed using Equation (7) showed no variation with heliographic latitude, Λ (as expected from the Ulysses result) but did increase with heliocentric distance, r, as shown in Figure 18. This rise means that there are two components of the flux values computed using Equation (6): the coronal source flux, F_S (which is the value of $F(r)$ for $r = 2.5\,R_\odot$) and a component which arises in the heliosphere (and grows with r). Note that Figure 18 shows that this rise is present at $r < r_1$ as well as $r > r_1$, such that ΔF is negative at $r < r_1$ for the definition given in Equation (7). This increase can arise from a number of physical phenomena, including Alfvén wave growth, transients such as CMEs, kinematic effects of solar wind flow structure on the frozen-in field, and the outward propagating structures, such as plasmoids and folded flux tubes, generated by near-Sun reconnection of flux. All these effects can give Toward and Away field structure in the heliosphere within a single coronal source sector (i.e., a region of constant unipolar radial field polarity at the source surface): and this heliospherically-imposed toward-and-away structure at general r has a range of characteristic spatial scales such that it will only partially be averaged out by taking the modulus of a mean over a given timescale T (Lockwood *et al.*, 2006a). Thus, ΔF varies with T and although most Alfvén waves are averaged out at relatively small T, the other phenomena are not (Lockwood and Owens, 2013). The problem

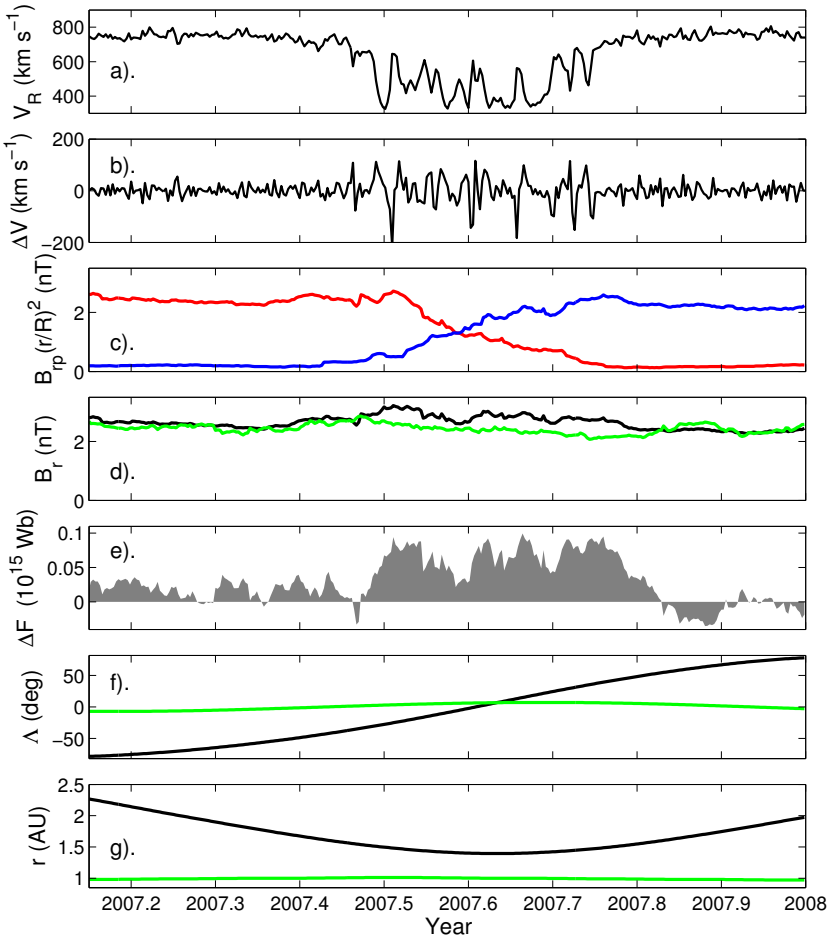

Figure 17: A comparison of Ulysses and ACE data during the third perihelion pass of Ulysses, which took place between day 36 of 2007 and day 13 of 2008. In all panels, black lines are data from Ulysses while green lines are from ACE, which is in a halo orbit around the L1 Lagrange point. From top to bottom: (a) daily means of the solar wind speed, $V_{\rm SW}$; (b) The difference between successive daily $V_{\rm SW}$ values, ΔV; (c) The 27-day running means of the away and toward (red and blue lines, respectively) radial field components B_{rp} observed at Ulysses and averaged separately; (d) 27-day running means of the normalized absolute radial field magnitude at Ulysses, $(r_U/r_1)^2\langle|B_{rU}|_T\rangle_{27\,\rm d}$ (black line), and ACE, $\langle|B_{rA}|_T\rangle_{27\,\rm d}$ (green line); (e) The flux excess at Ulysses, defined as $\Delta F = 2\pi\times[r_U^2\langle|B_{rU}|_T\rangle - r_1^2\langle|B_{rA}|_T\rangle_{27\,\rm d}]$ (grey shaded area); (f) The heliographic latitude of Ulysses (Λ_U, black line) and ACE (Λ_A, green line); (g) The heliocentric distance of Ulysses (r_U, black line) and ACE ($r_A = r_1$, green line). In all cases the radial field values are the asymptotic values for the averaging timescale T approaching zero. Image reproduced by permission from Lockwood *et al.* (2009b), copyright by AGU.

is that using too large a T results in toward-and-away field that does map back to the coronal source structure also being cancelled. (Taken to its extreme, use of $T = 27$ days would result in $F(r) \approx 0$ as toward and away sector flux cancel). There is no T at which the heliospheric effects are completely averaged out but none of the source sector structure is averaged out and so using a fixed T to eliminate this effect must be a form of compromise. Lockwood *et al.* (2009c) devised a method to use the measurements of the tangential field and flow to map from a general r to r_1 and thereby evaluate how much radial field structure has been amplified by longitudinal flow structure in the heliosphere. Such effects are known to take place, for example, prolonged intervals a near-radial heliospheric field (Jones *et al.*, 1998) have been explained by Riley and Gosling (2007) in terms of this effect. Because this was based on the frozen-in flux theorem and simple kinematics, Lockwood *et al.* (2009c) termed this the "kinematic correction". It is a simple, single correction that has to account for the variety of effects discussed above; however, because for all of these phenomena the frozen-in theorem applies, it does have the potential to make allowance for them. Short period Alfvén waves, and any other small-scale B_r polarity structure, on timescales shorter than 1 h was averaged out by using $T = 1$ h. The lines in Figure 18 give the distribution of the variations in ΔF predicted using the kinematic correction for each Carrington rotation from the mean variability in solar wind speed, ΔV: it can be seen that these fit the distribution of observed values very well.

Figure 19 compares the raw data shown in Figure 17 (for $T \rightarrow 0$ top panel) with results obtained using $T = 1$ day (middle panel) and using $T = 1$ h with the kinematic correction of Lockwood *et al.* (2009b) (bottom panel). In all cases the thick black line is the Ulysses data, normalised to $r = r_1$ by multiplying by $(r_U/r_1)^2$, and the thin line is from the simultaneous ACE data in the ecliptic plane. The middle panel shows that using an averaging timescale $T = 1$ day before taking the modulus has reduced all the ACE data, which is expected because ACE is within the streamer belt and observed both Toward and Away field which cancels to some degree when $T = 1$ day is used. The Ulysses values are also reduced in the same way, particularly while it is within the streamer belt, but also to a lesser extent while the spacecraft is within the polar coronal holes: hence although good agreement is now obtained when both craft are within the streamer belt, this is not the case when Ulysses is in the polar coronal holes. The third panel shows the effect of applying the kinematic correction computed from the solar wind variability ΔV using the equations of Lockwood *et al.* (2009b). With this correction, the Ulysses data are no longer showing enhanced values where ΔV is enhanced and good agreement is obtained between the Ulysses and ACE data throughout the pass.

As noted above, the open solar flux F_S is the value of $F(r)$ for $r = 2.5\,R_\odot$, inserting this into Equation (7) and re-arranging, we get an equation that allows us to use near-Earth IMF data to compute the open solar flux:

$$F_S = 2\pi r_1^2 \langle |B_r(r_1)|_T \rangle_{27\,\mathrm{d}} - \Delta F_S\,, \tag{8}$$

where ΔF_S is the excess flux between the coronal source surface and near-Earth space (defined by Equation (7) because $\Delta F = \Delta F_S$ for $r = 2.5\,R_\odot$). Lockwood and Owens (2009) surveyed all the Ulysses data and showed that using $T = 1$ h and the kinematic correction to evaluate ΔF_S reduced the overall error in F_S, computed using Equation (8), to just $\pm 2.5\%$.

As mentioned above, the PFSS method has also been used to compute open solar flux. In particular, Wang and Sheeley Jr (1995, 2003a) and Wang *et al.* (2000) used the Ulysses result to compare the value derived from near-Earth in-situ observations with PFSS-derived values and obtained good agreement. (Note that these authors actually compared $\langle |B_r|_T \rangle$ rather than F_S). However, two caveats to this comparison should be noted. First, the magnetogram data required re-processing using a latitude-dependent instrument saturation factor (Wang and Sheeley Jr, 1995) which has long been the subject of some debate (Svalgaard *et al.*, 1978; Ulrich, 1992; Riley *et al.*, 2013). Secondly, these authors used an averaging timescale of $T = 1$ day on the in-situ B_r data

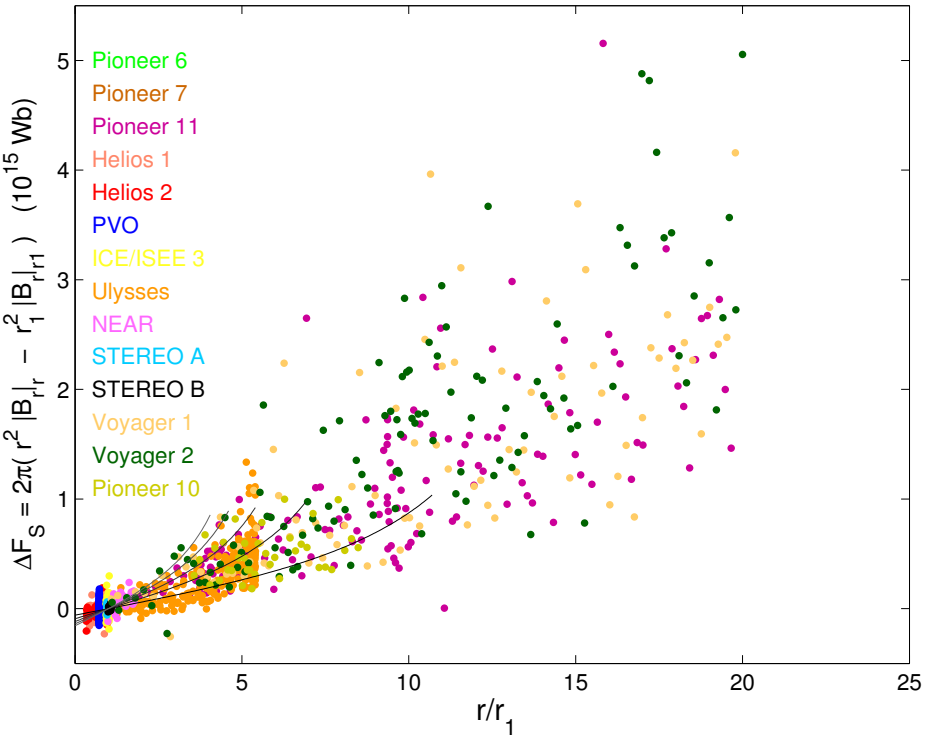

Figure 18: A comparison of ΔF values, as a function of r/r_1 (where $r_1 = 1\,\mathrm{AU}$), computed using Equation (7) for $T = 1$ h and averages over Carrington rotations (i.e., $\langle|B_r(r)|_{1\,\mathrm{h}}\rangle_{27\,\mathrm{d}}$ is used for both the heliospheric and simultaneous near-Earth data) from 14 heliospheric spacecraft: Pioneer 6, Pioneer 7, Pioneer 10, Pioneer 11, Helios 1, Helios 2, Pioneer Venus Orbiter (PVO), International Sun-Earth Explorer ISEE3 (later renamed International Cometary Explorer, ICE), Ulysses, Near-Earth Asteroid Rendezvous (NEAR), Solar Terrestrial Relations Observatory (STEREO) A, STEREO B, Voyager 1, and Voyager 2. The lines show the distribution of predicted ΔF from the "kinematic correction" calculated for each sample (Lockwood *et al.*, 2009c), which are exceeded a fraction p of the time, where p is 0.1 (for black line), 0.25, 0.5, 0.75, and 0.9 (light gray line). Image reproduced by permission from Lockwood *et al.* (2009c), copyright by AGU.

before taking the modulus. This means, effectively, that they used Equation (8) with $T = 1$ day, for which $\Delta F_S \approx 0$. The middle panel of Figure 19 shows that this works reasonably well within the streamer belt. Lockwood *et al.* (2006b) showed explicitly the effect of the value of T on this comparison and that $T = 1$ day is the optimum value on average. However, although it makes ΔF_S go to zero on average, it is an approximation. Effectively, it is chosen as a compromise that averages out much of the heliospherically-generated component in $\langle|B_r|_T\rangle$, without averaging out too much of the source sector structure.

Figure 20 shows that the difference between the PFSS-derived values of open solar flux by Wang and Sheeley Jr (1995, 2003a) and the mean value of $F(r_1)$ (for $T = 1$ h) is consistent with the rise in $F(r)$ values with r found using a wide variety of spacecraft in the heliosphere, as found by Owens *et al.* (2008a) (also using $T = 1$ h) and shown in Figure 18. Hence the PFSS data also strongly support the kinematic correction. The degree to which the PFSS data and the near-Earth data are consistent (if the kinematic correction is deployed) is underlined by the variation of derived annual means shown in Figure 21. The agreement is very good and better than using the $T = 1$ d, $\Delta F_S = 0$ approximation.

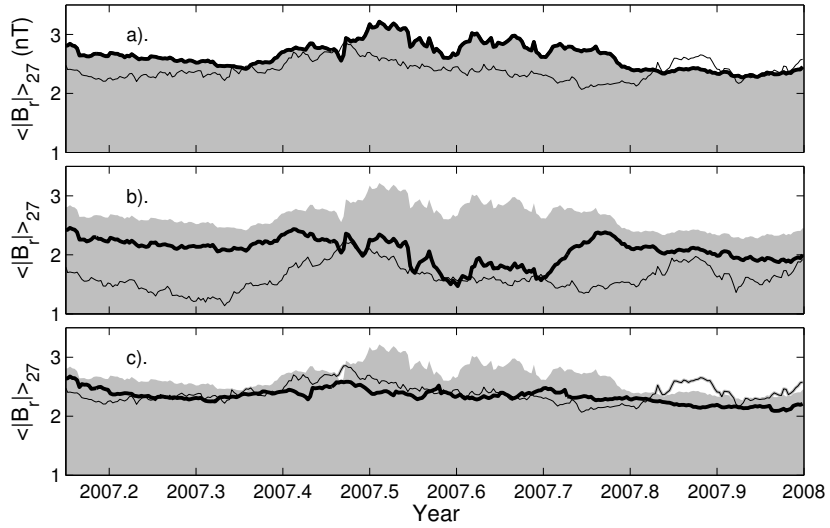

Figure 19: Application of the "kinematic correction" to the third Ulysses perihelion pass shown in Figure 17. All three panels show 27-day running means of various estimates of the absolute radial field (normalized to $r = r_1 = 1$ AU) from measurements by Ulysses (thick black lines) and ACE (thin black lines). Ulysses data from the top panel are shown by the area shaded grey in all three panels to facilitate comparisons. a) the asymptotic limit of the timescale $T \to 0$, $\langle |B_r(r_1)|_{T\to 0}\rangle_{27\,\mathrm{d}}$ (for ACE data, thin black line) and $(r_U/r_1)^2\langle |B_r(r_U)|_{T\to 0}\rangle_{27\,\mathrm{d}}$ (for Ulysses, thick black line); (b) the same for $T = 1$ day, $\langle |B_r(r_1)|_{1\,\mathrm{d}}\rangle_{27\,\mathrm{d}}$ (thin black line) and $(r_U/r_1)^2\langle |B_r(r_U)|_{1\,\mathrm{d}}\rangle_{27\,\mathrm{d}}$ (thick black line); and (c) data for $T = 1$ h with the kinematic correction applied to the Ulysses data to allow for heliospheric effects between r_U and r_1, $\langle |B_r(r_1)|_{1\,\mathrm{h}}\rangle_{27\,\mathrm{d}}$ (thin line) and $(r_U/r_1)^2\langle |B_r(r_U)|_{1\,\mathrm{h}}\rangle_{27\,\mathrm{d}} - \Delta F/(2\pi r_U^2)$ (thick black line). Image reproduced by permission from Lockwood *et al.* (2009b), copyright by AGU.

7.5 The use of the modulus of the radial field

Smith (2011, 2013) argues that the use of the modulus of the radial field B_r causes the excess flux and that the use of the kinematic correction is unnecessary. The first point is a somewhat semantic one but does have some validity. However, the second point does not follow from the first because it assumes that there is a problem-free and error-free alternative, which is not true. In fact the word "causes" is somewhat misleading: what taking the modulus (after averaging over an interval T) really does is cancel opposite polarity B_r within the interval T. The alternative advocated by Smith (2011, 2013) is that B_r be averaged over the durations of toward and away source field sectors. This is indeed, in principle, absolutely correct; however, it pre-supposes that in implementation one knows where the source sector boundaries lie. We here call this method "the variable-T method", because it is identical to employing the modulus, except that the interval durations used T are not fixed but are varied to cover the source sector durations. It is important to stress here that it is the sector boundaries at the coronal source surface, not at the observing spacecraft, that matter: Lockwood and Owens (2013) show that if the polarity changes at the spacecraft are used to define the sectors then the result is mathematically identical to that obtained by taking the modulus. Therefore, to implement the variable-T method, decisions are required about where the source sector boundaries are and where they map to on the satellite orbit, and what is a genuine source sector rather than heliospherically-imposed B_r polarity structure.

It is not adequate or acceptable to sidestep this issue by saying that source sector boundary crossings can be readily identified. For example, using electrons with energies > 2 keV to determine the connectivity between near-Earth space and the source surface, Kahler *et al.* (1996) found clear

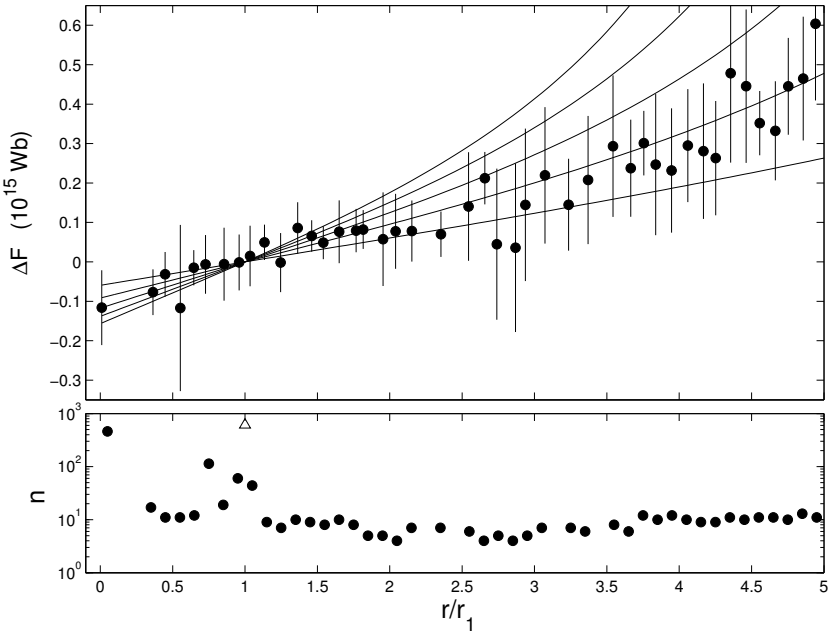

Figure 20: *Top:* The excess flux ΔF from Equation (8) as a function of r/r_1. The data points show the mean values from the spacecraft data shown in Figure 18, averaged over bins in r that are 0.1 AU wide, with error bars of plus and minus one standard deviation. The lines are the same as shown in Figure 18. The data point at the lowest r/r_1 is the entire PFSS data set for the coronal source surface at $r = 2.5\,R_\odot$. *Bottom:* The number of full Carrington rotation averages n contributing to these means. The open triangle shows the number in the OMNI-2 data set from $r = r_1$. Note that n is shown on a logarithmic scale. Image reproduced by permission from Lockwood *et al.* (2009c), copyright by AGU.

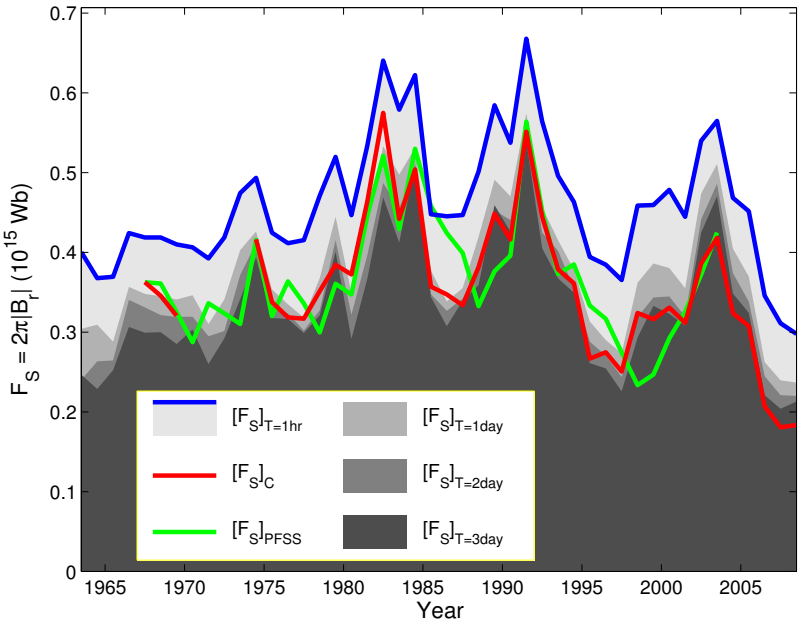

Figure 21: The variation of various open solar flux estimates during the space age. The gray-shaded areas show annual means of the signed open solar flux from the OMNI-2 data derived using $\Delta F_S = 0$ and the absolute value of the radial field taken of means on timescale T: the light gray area bounded by the blue line is for $T = 1$ h, and successively darker gray areas are for $T = 1$, 2, and 3 days. The green line is the variation from solar magnetograms using the PFSS method. The red line shows the OMNI-2 values for $T = 1$ h, minus the correction term for kinematic effects, ΔF_S . Image reproduced by permission from Lockwood *et al.* (2009c), copyright by AGU.

examples of opposite polarity field within well-defined source sectors and also found that in all cases they showed bidirectional electron streaming which they associated with CMEs on a wide range of spatial scales. For CMEs within an inferred source sector, the flux would be cancelled out by using the variable-T method, but would thread the source surface and hence CMEs are one source of potential error. The net heat flux from the electron data show regions where the heat flux is towards the Sun. These unambiguously reveal "folded flux" (in which source toward/away field is folded so it points away from/toward the Sun at greater r (e.g., Owens *et al.*, 2013). Folded flux revealed by the suprathermal electron flows is often found in the vicinity of sector boundaries (Kahler *et al.*, 1998; Crooker *et al.*, 2004) along with seemingly plasmoidal structures (Foullon *et al.*, 2011) (which may, in some cases, actually be folded flux that has latitudinal structure). In addition, Owens *et al.* (2013) have recently shown folded flux is also associated with pseudostreamers embedded within IMF polarity sectors. As a result of folded flux, Kahler and Lin (1995) deduced that some sector boundaries did not show local field reversals at the spacecraft and some field reversals were seen away from the true sector boundaries. Thus defining where the source sector crossing point is from the associated polarity change in B_r at the spacecraft is not straightforward and would contribute to an unknown error to the derived open solar flux.

Thus, for the derivations of open solar flux from interplanetary craft to be repeatable, a full catalogue of assumed source sector crossings would be needed for the variable-T method and the error introduced into open solar flux estimates by inaccuracies in that catalogue would not be known. On the other hand, using the modulus with kinematic correction (or the easier-to-implement $T = 1$ day method) is a repeatable algorithm that does not depend on a catalogue

of sector crossings. Another advantage of the modulus approach is that it makes the issue of heliospherically-produced radial field structure explicit and does not hide it in the choice of the intervals to average over in the variable-T method.

Lockwood and Owens (2013) argue that the variable-T and modulus approaches are actually complementary, each with its own strengths and weaknesses, and that it is neither correct nor helpful to advocate the use of one over the other: rather, it is important to fully understand the strengths, limitations and applicability of both. What is most interesting is that the two approaches generate similar answers in several important respects (for example, both give the Ulysses result of the latitudinal constancy of the radial field.) In addition, Lockwood and Owens (2013) point out that both methods average along a sector of the spacecraft orbit and so both are subject to errors (that are different in the two cases) associated with latitudinal variations in folded flux tubes.

It is here worth recording that there is another approach discussed in the literature which allows values of F_S from PFSS to be matched to the values derived from in-situ spacecraft measurements. Erdős and Balogh (2012) analysed the field in a reference frame aligned to the Parker spiral direction, as computed from the frozen-in flux theorem using the observed solar wind flow speed (see Equation 5). Like the fixed-T (modulus) methods it has the advantage of being repeatable without implementing a catalogue of source sector boundary crossings. However, it should be noted that disconnected and folded flux is often well-aligned with the Parker spiral (Crooker *et al.*, 2004; Foullon *et al.*, 2011; Owens *et al.*, 2013) and that the authors used an averaging interval of $T = 6$ h to match the open solar flux estimates derived from the satellite to those from the PFSS values, and hence there is an element of fixed-T averaging in this method also.

8 Regression Techniques

Because the reconstructions all rely on finding relationships between modern geomagnetic activity data and simultaneous measurements of near-Earth space, regression techniques are needed to enable extrapolation to before the space age. The discussion in the literature between Svalgaard and Cliver (2005), Lockwood *et al.* (2006a), and Svalgaard and Cliver (2006) highlights the many pitfalls in this area.[1] This discussion was on the use of the IDV index. Svalgaard and Cliver (2005) (SC05) used IDV to conclude that B increased by only 25% between the 1900s to the 1950s and that this was in contrast to the more than doubling of B which they argued was inherent in the results of Lockwood *et al.* (1999a). Lockwood *et al.* (2006a) (LEA06) pointed out that some of this difference was due to the fact that Lockwood *et al.* (1999a) actually reported a doubling in the open solar flux, F_S, not B (as shown later by Figure 29, F_S is not proportional to B). However, there were several other factors, all of which worked in the same direction and so combined to make the estimate of the drift by SC05 exceptionally low. SC05 employed a simple ordinary linear least squares (OLS) regression which yielded residuals that showed heteroscedasticity, some non-linearity, and a systematic bias and which do not have a Gaussian distribution, thereby violating central assumptions of least-squares regression and showing the derived fit is unreliable. The regression results of SC05 were strongly influenced by outliers, which applied great leverage to their regression fit. More reliable regressions were obtained by LEA06 using least median squares (LMS) regression and, better still, using Bayesian statistics (the BLS procedure employed by REA07). In addition, SC05 attempted to fill in the data gaps in the IMF data using a 27-day recurrence technique, despite the relatively low autocorrelation functions of the IMF at 27 day lags, and LEA06 show that this also caused a slight underestimation of the long-term drift (piecewise removal of the IDV data during IMF data gaps is much more reliable). Lastly SC05 under-estimated the long-term change in their own results.

The initial response by Svalgaard and Cliver (2006) did not accept these arguments, but as shown in the following sections, a subsequent reconstruction by Svalgaard and Cliver (2010) is in very good agreement with the Lockwood *et al.* (1999a) reconstruction. In fact, the change between the Svalgaard and Cliver (2010) and SC05 reconstructions of IMF B was almost exactly what was called for by the residuals analysis of LEA06. This change was caused by the availability of just four additional annual mean datapoints near the long and low minimum between cycles 23 and 24 for which B was low. The fact that change was needed in response to the addition of just a few more datapoints confirms that the original SC05 fit was not robust.

Because the potential pitfalls in regression techniques can have such a major effect on the reconstructions, it is worth exploring the relative merits of the various linear regression procedures used in this context. Figure 22 stresses how much they can differ, showing the scatter plot and the various regressions between annual means of the IDV index and the IMF, B. SC05 used OLS but the slope they derived is slightly lower than LEA06's implementation of OLS because of their different treatment of data gaps. OLS gives the lowest slope, whereas BLS gives the largest.

The details of the regression procedures (with appropriate references for the statistical techniques) and discussion of their relative merits and pitfalls are given in the paper by LEA06. The advantage of the LMS procedure is that it is not as influenced by outliers that can change the slope of the fit dramatically if they have a high value of the Cook-D leverage factor. The MAA (Major Axis Analysis) procedure is inappropriate in the context of these reconstructions and the BLS procedure is as employed by Rouillard *et al.* (2007) (REA07). The tests described below show that BLS performed best.

Notably, the OLS procedure used by SC05 gives lowest slope in Figure 22, and so would give

[1] In the following: Lockwood *et al.* (1999a) – LEA99; Lockwood *et al.* (2006a) – LEA06; Lockwood *et al.* (2009d) – LEA09; Rouillard *et al.* (2007) – REA07; Steinhilber *et al.* (2010) – SEA10; Svalgaard and Cliver (2005) – SC05; Svalgaard and Cliver (2007a) – SC07; Svalgaard and Cliver (2010) – SC10.

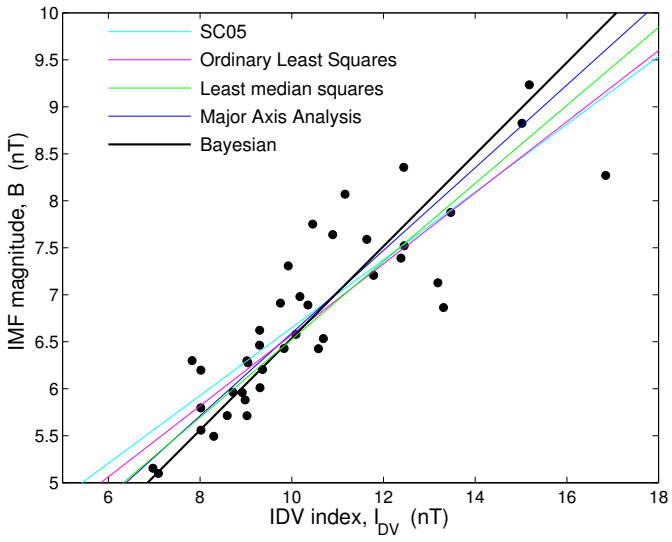

Figure 22: Scatter plot and best-fit linear regressions between the IDV index and the IMF strength B. The wide variety of regression slopes demonstrates the effect of the choice of regression method used. The light blue line is the fit by SC05, the other lines are the best-fit regressions by LEA06 for residual minimization methods of: (mauve line) Ordinary Least Squares (OLS); (green line) least median of squares (LMS); (blue line) Major Axis Analysis (MAA); (black line) Bayesian Least Squares (BLS).

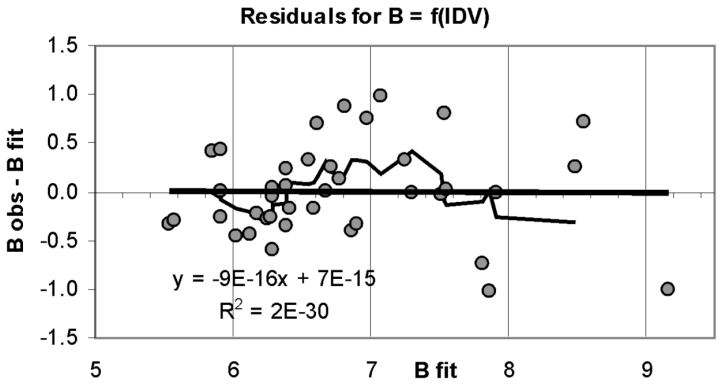

Figure 23: Analysis by Svalgaard and Cliver (2006) of the fit residuals in the fit of IDV against B by Svalgaard and Cliver (2005). The fit residuals are plotted as a function of the fitted value, which is the correct test for homoscedasticity. Image reproduced by permission from Svalgaard and Cliver (2006), copyright by AGU.

the lowest long-term drift in the reconstruction of open solar flux. There are a number of ways of evaluating the quality of a regression fit. One of the most important is to check that the fit residuals are randomly and normally distributed: the fit of IDV against B used by SC05 is analysed in Figures 23, 24, and 25.

Fits should be homoscedastic, i.e., the residuals should not show a trend in their spread. In their reply, Svalgaard and Cliver (2006) quite rightly state that this should be tested by plotting fit residuals against the fitted values. Figure 23 shows this residual plot which they claim shows the fit is homoscedastic because the mean of the residuals does not change with the fitted value. However, homoscedasticity requires the spread of fit residuals (not their mean value) does not change with the fitted value, and Figure 23 shows the spread does increase with increasing fitted values. Hence the fit is heteroscedastic rather than homoscedastic. In addition, the plot shows a marked tendency towards an inverted-U form which is characteristic of some nonlinearity.

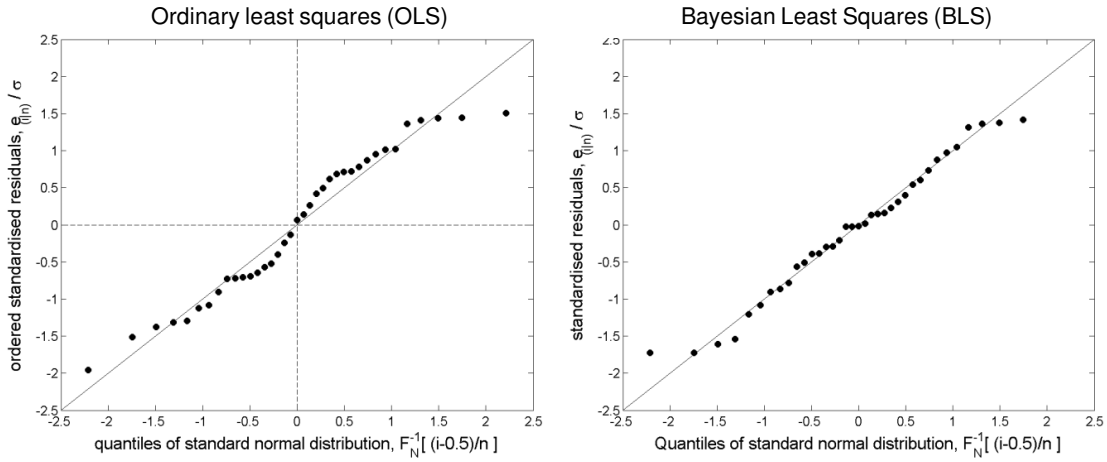

Figure 24: Quantile-quantile (QQ) plots for (left) OLS and (right) BLS regressions of the IDV index and the IMF strength B. These plots are used to check if the distribution of fit residuals is Gaussian, as assumed by least-squares fitting procedures. The ordered, standardized residuals $\Delta B_i/\sigma$ (where $\sigma^2 = \Sigma_i \Delta B_i^2/(n-2)$ and n is the number of samples) are shown as a function of the corresponding quantiles of a standard normal distribution. Deviations from the line of slope 1 shown reveal departures from a normal distribution of standard distribution σ.

A second test used by LEA06 was quantile-quantile (QQ) plots to test if the residuals were normally distributed, as is assumed by all least squares regressions. The standardized residuals are placed in order by size and plotted against the quantiles for a standard normal distribution. The deviations from the straight line of slope 1 reveal departures from a normal distribution. Figure 24 shows that the OLS fit gave considerably larger deviations from a Gaussian distribution of residuals than did the BLS fit and so the BLS method is giving the more valid least-squares regression.

Because they found the SC05 fit was heteroscedastic, potentially nonlinear and failed the QQ test for normally-distributed residuals, LEA06 tested for a trend in the residuals by plotting the fit residuals as a function of the observed values. The results are shown in Figure 23 for three regression prcodures. The BLS regression meets the requirement that there is no trend (and LEA06 show the LMS regression does as well) but the MAA and OLS fail this bias test. They underestimate the trend in the data because $B_{\rm fit}$ is consistently an overestimate of $B_{\rm obs}$ when $B_{\rm obs}$ is small and consistently an underestimate of $B_{\rm obs}$ when it is large. Thus, reconstructions based on this OLS fit will self-evidently underestimate the true range of variation in IMF B.

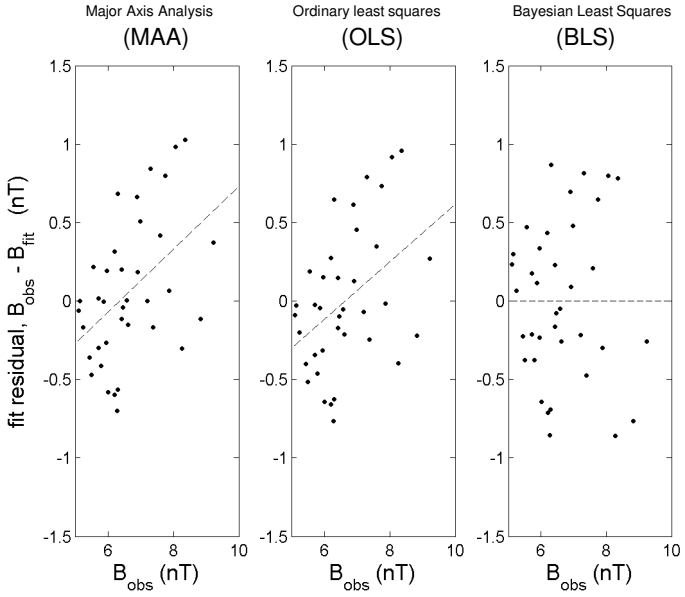

Figure 25: Analysis of the fit residuals in the fit of IDV against B using: (left) MAA; (middle) OLS; and (right) (BLS). The fit residuals ($B_{\rm obs} - B_{\rm fit}$) are shown as a function of $B_{\rm obs}$, where $B_{\rm obs}$ is the observed value and $B_{\rm fit}$ is the fitted value from the IDV index. The dashed line is the linear regression fit to the points to highlight trends. The OLS and MAA fit residuals show strong trends and so should not be used because $B_{\rm fit}$ is consistently an overestimate of $B_{\rm obs}$ when $B_{\rm obs}$ is small and consistently an underestimate of $B_{\rm obs}$ when it is large. Hence, these fits seriously underestimate the real trend in the data. In contrast, the BLS regression is free of such a trend.

The better the correlation between two parameters, the more similar will be the results of the various regression procedures and, hence, these tests would become increasingly less important. Given that small changes in the fitted slope will make very large differences to the maximum and minimum values seen in a reconstruction, it is very important that these tests are carried out to ensure that the optimum regression procedure has been used and any one regression fit is valid.

9 Reconstructions

The previous sections give the physical principles that allow us to reconstruct annual means of heliospheric parameters from geomagnetic activity. By using a combination of geomagnetic activity indices that have differing sensitivities to the substorm phenomenon (for example, one range index and one based on interdiurnal variability can be used) both the IMF magnitude B and the solar wind speed, $V_{\rm SW}$, can be derived. From Parker spiral theory the modulus of the radial field $|B_r|$ can be deduced and using the Ulysses result, this means the open solar flux, F_S can be reconstructed too. The first paper to exploit this possibility was Lockwood *et al.* (1999a) who reconstructed the open solar flux, F_S and found considerable variation with 11-year running means around 1985 more than twice those found around 1900. A wide variety of different procedures have subsequently been used. The most obvious differences between them are the basis geomagnetic data employed. However, there are other differences. In this section we look at the resulting reconstructions of B, $V_{\rm SW}$, and F_S.

9.1 Results for the near-Earth IMF

Figure 26 shows the reconstructions of annual means the IMF magnitude, B, and compares them to annual means of the near-Earth measurements. The reconstructions by Lockwood *et al.* (2009d) (LEA09, red line) and Rouillard *et al.* (2007) (REA07, blue line) are based on the aa_C (range) index and the m index. The Svalgaard and Cliver (2010) reconstruction (SC10, thin black line) is based on the IDV index alone. Because the method of SC10 is relatively straightforward, making use of a single correlation, it has the advantage that it is much easier to study the propagation of uncertainties and the grey area shows the error in the SC10 reconstruction, as evaluated by Lockwood and Owens (2011) from the regression error estimates by Svalgaard and Cliver (2010). These estimates are made using the standard equations for uncertainty in the slope and intercept of a linear regressions which are approximate (Richter, 1995) and do not allow for experimental uncertainties in the data used. A more rigorous examination of uncertainty is presented in Section 9.4.2. The orange line (LEA99) was not published in the paper by Lockwood *et al.* (1999a), which focussed on open solar flux. However, Lockwood and Owens (2011) have extended the procedure of Lockwood *et al.* (1999a) to evaluate B, the results of which is shown here. In this procedure, only the aa_C index was used, and the effect of $V_{\rm SW}$ on this range index was allowed for using the recurrence index of aa_C (based on the autocorrelation of aa_C at 27-day lag). This works because in the streamer belt, annual means of $V_{\rm SW}$ are enhanced by the intersection with fast solar wind streams emanating from isolated low-latitude coronal holes or from low-latitude extensions to polar coronal holes (Sheeley Jr *et al.*, 1976; Wang and Sheeley Jr, 1990). These fast streams interact with the slow solar wind ahead of them and, because the coronal holes generally persist for several solar rotations, cause co-rotating interaction regions which give recurrent geomagnetic disturbances, so increasing the aa_C recurrence index. As a result, recurrence indices (e.g., Sargent, 1986) are correlated with the average solar wind speed.

It can be seen that, despite the wide variety of implementations and the variety of different geomagnetic data used, the reconstructions of B agree remarkably well and all agree well with the observed values from the space age. As expected, the differences get somewhat larger as one goes back in time. The year 1901 appears to either contain some error in one or more of the data series that is propagated into some of the reconstructions, or has exposed a limitation in one of the procedures. The paper by Rouillard *et al.* (2007) offers some insight into this as they applied both ordinary least-squares regression (OLS) and Bayesian least-square regression (BLS) techniques to two combinations of indices: aa_C with m and aa_C with IDV. Their Figure 5 reveals that a shallow minimum is found for 1901 using the aa_C with IDV pairing with OLS. The same pairing gives a somewhat deeper minimum if BLS is used and the variation shown in Figure 26 is

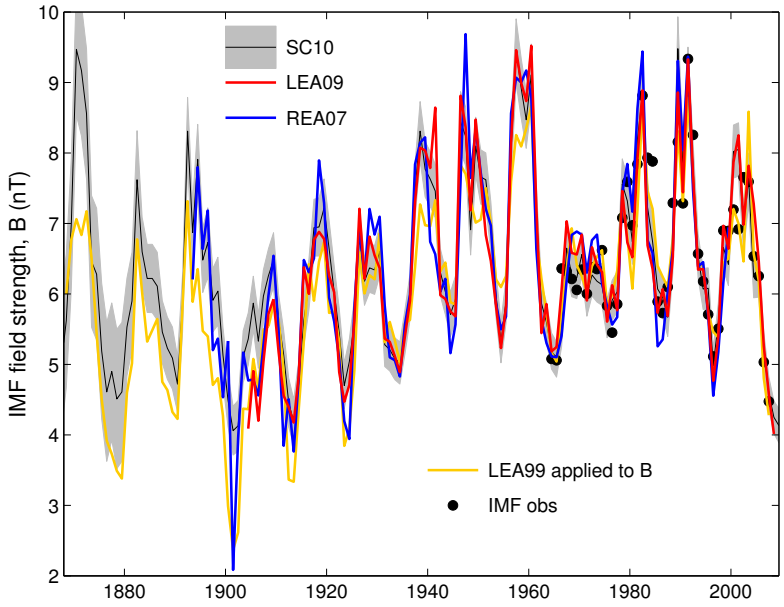

Figure 26: Reconstructions of annual means of the near-Earth IMF field strength B from geomagnetic activity data: LEA99 (orange) uses the method of Lockwood *et al.* (1999a) applied to aa_C; LEA09 (red) is by Lockwood *et al.* (2009d) and uses aa_C and m; REA07 (blue) is by Rouillard *et al.* (2007) and also uses aa_C and m; SC10 (thin black line with uncertainty shown by the surrounding grey area) is by Svalgaard and Cliver (2010) and uses IDV. Solid circles show annual means of IMF observations from the OMNI2 database. Note that Lockwood *et al.* (1999a) did not reconstruct B and the LEA99 variation shown was generated by Lockwood and Owens (2011) who adapted the Lockwood *et al.* (1999a) procedure to predict B. Image reproduced by permission from Lockwood and Owens (2011), copyright by AGU.

found using the aa_C with m pairing and, again, is somewhat deeper if BLS rather than OLS is used. The figure shows that using aa_C alone gave a lower value but Figure 33 shows the $IDV(1d)$ index with a non-linear fit and full error analysis (Lockwood *et al.*, 2013b) gives a similar shallower 1901 minimum to that from an OLS fit to IDV (Svalgaard and Cliver, 2010). In itself the larger spread of estimates for this one year (1901) is not significant, other than it does highlight that uncertainties are larger when the number of available stations is lower.

9.2 Results for the near-Earth solar wind speed

The SC10 and LEA99 papers do not reconstruct the solar wind speed variation but Svalgaard *et al.* (2003), Svalgaard and Cliver (2007a) and REA07 did. Furthermore, REA07 used a number of different procedures to check how robust their reconstructions are: their results for V_{SW} are shown in Figure 27. Specifically they employed two different regression procedures: ordinary least squares (OLS) and Bayesian Least Squared (BLS). They also used two different combinations of geomagnetic indices: aa_C with m and aa_C with IDV, which as illustrated Figures 3 and 16, have differing dependencies on the solar wind speed.

The derived variations in annual means of V_{SW} are very similar in all four cases, as is the Svalgaard and Cliver (2007a) reconstruction (not shown). All show a weak upward trend on average during the past century, but this trend is not as strong as that in the IMF B. This agrees with recent inferences from Wang and Sheeley Jr (2012) that the lower IMF in low-activity cycles would be accompanied by lower solar wind number density, but not significantly lower solar wind

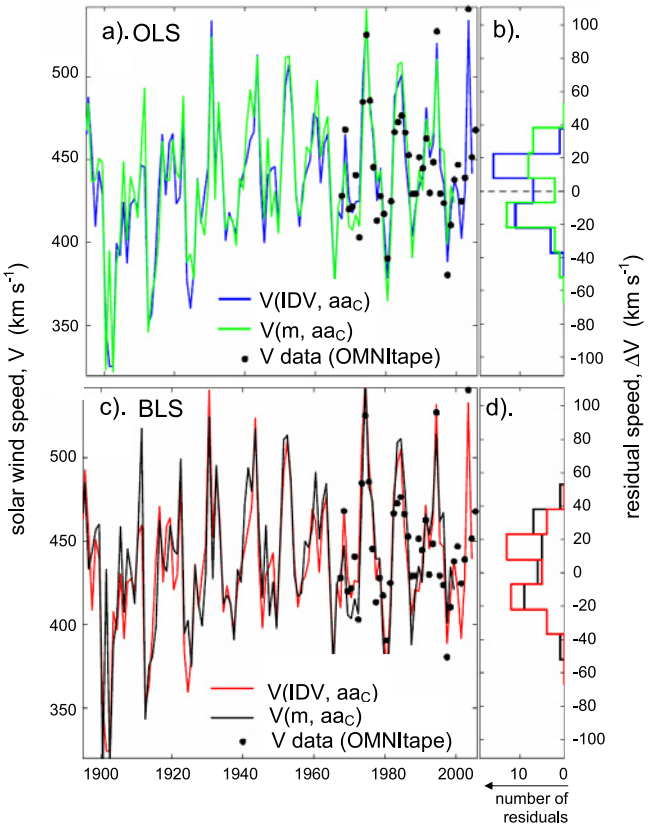

Figure 27: Reconstructions of annual means of the near-Earth solar wind speed V_{SW} from geomagnetic activity data. The upper panel is for ordinary least squares (OLS) regression, the lower panel for Bayesian least squares regression (BLS). Plots to the left are the time series, those to the right give the distribution of residuals of the fit to interplanetary in-situ observations. Blue and red lines are derived using aa_C and IDV whereas green and black are using a_C and m. Solid circles show annual means of solar wind speed observations from the OMNI2 database. Image reproduced by permission from Rouillard *et al.* (2007), copyright by AGU.

speed. Peaks in V_{SW} are seen in the declining phase of the solar cycles. In the space age there has been a marked tendency for these to be larger for even numbered cycles than odd-numbered ones (Hapgood, 1993) but this is not seen in the reconstruction for before the space age. This agrees with the inference from the 27-day recurrence in the *aa* index (see the lower panel of Figure 1 of Lockwood *et al.*, 1999a). As yet we have no explanation for this difference between even- and odd-numbered cycles, nor why it appears to be intermittent on centennial timescales.

9.3 Results for the open solar flux

Figure 28 shows the open solar flux reconstructions corresponding to the IMF reconstructions shown in Figure 26. Note that Svalgaard and Cliver (2010) did not compute F_S because the main focus of their paper was the IMF B, but their reconstruction (and the uncertainty band around it) has here been converted into open solar flux using the polynomial fit in Figure 29 (see below). Again the agreement is generally good, but larger differences do exist than for the IMF reconstructions. In particular, the original reconstruction by Lockwood *et al.* (1999a) (LEA99,

the orange line in Figure 28) was derived using the aa index and hourly mean data of the IMF, with no kinematic correction to the $|B_r|_{1\,\mathrm{h}}$ values. It can be seen this gives larger values than the best open solar flux values from in-situ data which do deploy the kinematic correction (the black dots in Figure 28). The green line in the figure (labelled LEA_C) shows the results of applying the LEA99 procedure to the aa_C index and using kinematically-corrected IMF B_r values. Comparing the green and orange lines it can be seen that applying these corrections has lowered the open solar flux estimates at all times, but the effect is greatest for the modern data (after 1957). Before 1957 (which is when the move of the northern hemisphere aa station from Abinger to Hartland generates the major difference between aa_C and aa) the difference between the two is not as great. As in Figure 26, the red and blue lines are from Lockwood *et al.* (2009d) (LEA09) and Rouillard *et al.* (2007) (REA07). It can be seen that the LEA99 procedure, when applied to the same data as used by LEA09 and REA07 generates very similar results, despite being based on aa_C and the 27-day aa_C recurrence index, whereas LEA09 and REA07 are based on combining aa_C and m. Given it is based on their B, it is not surprising that the SC10 reconstruction is, as for B, slightly larger than the others in the earliest years; nevertheless, agreement is remarkably close overall.

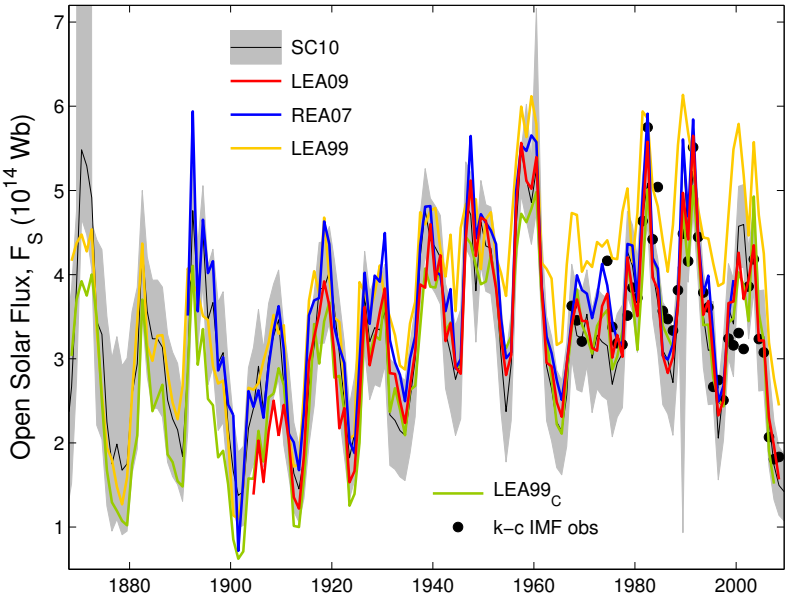

Figure 28: Reconstructions of annual means of the open solar flux F_S from geomagnetic activity data: LEA99 (orange), LEA09 (red), REA07 (blue), SC10 (thin black line with uncertainty shown by the surrounding grey area). See text for details. The variations are as they appear in the publications except that attributed to SC10, which is their variation in B, converted to F_S using the polynomial fit shown in Figure 29. The green line ($\mathrm{LEA99}_C$) is derived from the LEA procedure, applied to the aa_C index and to interplanetary data with the kinematic correction applied to $|B_r|_{1\,\mathrm{h}}$. The solid circles are the values derived from interplanetary observations using the kinematic correction, described by Lockwood *et al.* (2009c).

Figure 29 shows the variation of F_S with B used to convert the SC10 data. This is a scatter plot of the data from the LEA09 reconstruction for 1905 – 2009 (black dots) and from the in-situ spacecraft data (open triangles). The black line is a polynomial fit (given by equation 8 of Lockwood and Owens, 2011), constrained to pass through the origin because if the open solar flux ever fell to zero, the near-Earth IMF would necessarily also fall to zero. This fit varies considerably from the best fit linear regression, shown by the dot-dash line. The form of the polynomial fit is readily understood in terms of the competition between two effects (Lockwood *et al.*, 2009d). The

first is what would be seen for uniform solar wind flow (over a 27-day period), as predicted by Parker spiral theory. Sections 9.1 and 9.2 show that as the average IMF B rose over the past 150 years, the average solar wind speed V_{SW} also rose slightly. This causes the spiral field to unwind such that the ratio $|B_r|/B$ rises and, hence, the ratio $|F_S|/B$ also rises as B rises. This is consistent with the sense of the curvature in non-linear behavior seen in the data and the polynomial fit at B below about 6 nT. However, at B above 6 nT, the ratio $|F_S|/B$ falls slightly as B continues to increase. This is consistent with an increased kinematic effect due to increased longitudinal structure in the solar wind at higher solar activity which will increase the B at $r = 1$ AU for a given F_S.

There is a point that has caused some confusion and needs clarifying here. The non-linearity of open solar flux F_S and near-Earth IMF B means that the radial component of the near-Earth IMF B_r is not linearly related to B. The original reconstruction of F_S by Lockwood *et al.* (1999a) used not only a linear relationship but proportionality between B_r and B (by assuming that on annual mean timescales the gardenhose angle was constant) and so approximated the data in Figure 29 with $F_S = sB$. However, this approximation was used to derive an analytic form for F_S that was then fitted to the *aa* and Lockwood and Owens (2011) have shown that, although this influences the fit coefficients, it does not greatly alter the derived F_S. In other words, $F_S = sB$ was a reasonable approximation to make in this context. However, Figure 29 shows that proportionality is an approximation and cannot be relied upon in general.

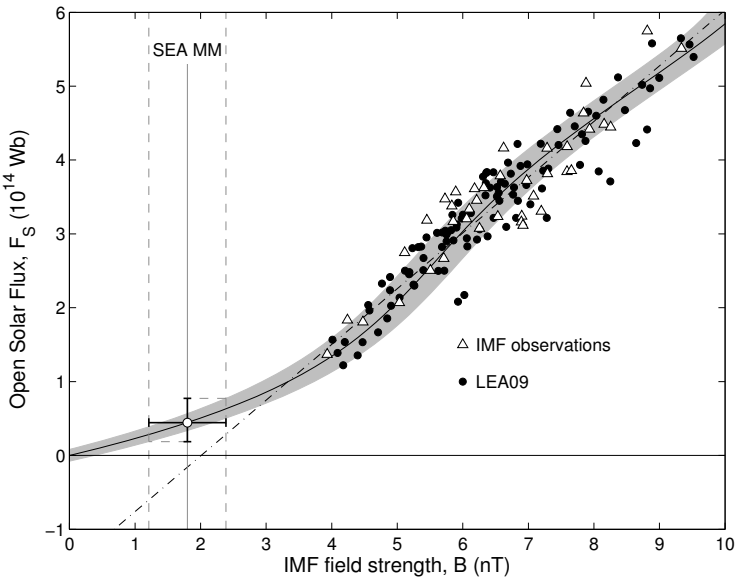

Figure 29: Scatter plot of open solar flux F_S as a function of near-Earth IMF field strength, B. The solid points are from the LEA09 reconstruction and the open triangles are annual means from in-situ interplanetary data (with F_S computed using Equation (8) with $T = 1$ h and the kinematic correction for ΔF_S). The black line is a polynomial fit to both datasets, constrained to pass through the origin $(B, F_S) = (0, 0)$, with an uncertainty given by the grey area. The dot-dash line is a linear regression fit. The vertical solid line labelled SEA MM shows the value of B during the Maunder minimum derived using cosmogenic isotopes by Steinhilber *et al.* (2010) and the vertical dashed lines bound the uncertainty in that estimate. Image reproduced by permission from Lockwood and Owens (2011), copyright by AGU.

Thus, there is considerable agreement between the various reconstructions of both the open solar flux and the near-Earth interplanetary field. The main difference is that the SC10 reconstruction gives slightly but persistently higher values in the early years, but we should expect agreement to

be less good at these times as the number of stations available, and the long-term stability of their instrumentation is necessarily lower for the early data. SC10 extend their sequence back to 1835, just three years after the establishment of the first geomagnetic observatory: discussion of the validity of this extension is presented in Section 9.4.2. The LEA99 reconstruction extends back to 1868 because they used the *aa* index only and this is the date at which Mayaud began his analysis of range at two antipodal stations. This sequence has been extended back to 1844 using data from a single station (Helsinki) by Nevanlinna and Kataja (1993) and Nevanlinna (2004); Lockwood (2003) used this to extend the open solar flux back to this date. LEA09 are more conservative in that they used the *m* index which only uses datasets and composites that extend into the era of space measurements and they argued that the hourly mean (or hourly spot value) data that meet this criterion too few and of insufficient accuracy before 1902 (for example giving the uncertainty in 1901). There is considerable (if not complete) agreement after 1901 and so this gives more than 100 years of reconstruction that can be used to train and evaluate models of the long term variation in the IMF and open solar flux, and these are discussed in Section 11. These models are based on the longest series of as-it-happened observations available to us, which is of sunspot number (see the Living Review by Hathaway, 2010).

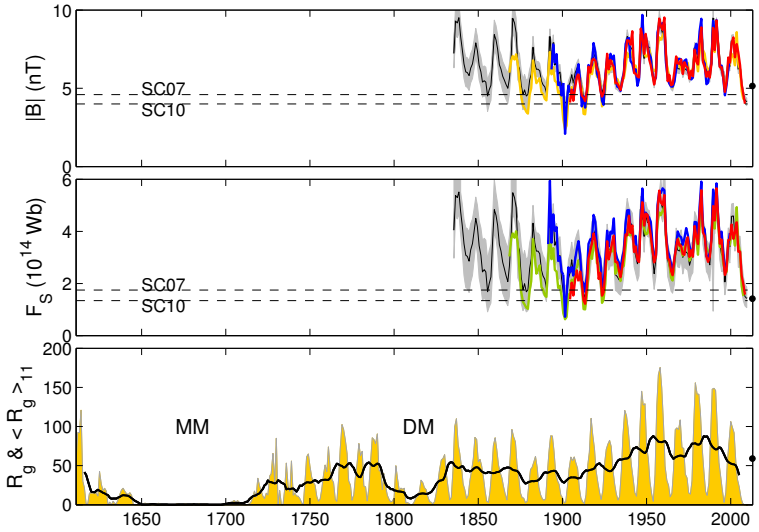

Figure 30: Long-term variations in (top) reconstructed IMF, B; (middle) open solar flux, F_S; and (bottom) group sunspot number, R_g. The upper two panels are as in Figures 26 and 28. In the bottom panel the orange-shaded area gives annual means of the group sunspot number and the black line their 11-year running means. The black dot to the right of each panel is the maximum 12-month mean seen thus far in cycle 24 at the time of writing (31 March 2013).

9.4 Discussion and uncertainty analysis

9.4.1 Comparison of reconstructions and the concept of "floor" values

Figure 30 compares the open solar flux reconstructions to the group sunspot number, R_g. The R_g data sequence was initially compiled by Hoyt and Schatten (1998). A number of possible adjustments to this sequence have been proposed recently, based on newly discovered historic observations. Reviewing these, after a consensus view has been reached, will be an important update to this and other living reviews. However, one adjustment, by Vaquero *et al.* (2011) is

already included in Figure 30 as this makes the decline into Maunder minimum conditions more consistent with cosmogenic isotope data (Lockwood *et al.*, 2011c). The upper dashed line in the top panel shows the "floor" in annual means of the near-Earth IMF B, of 4.6 nT postulated by Svalgaard and Cliver (2007b) (SC07). In this context, the author believes it is important to make a clear distinction between a genuine floor value (set by mechanisms which prevent the value of a given parameter from going any lower) and the minimum value detected since a certain date. The point is that without firm and quantitative understanding of the postulated mechanisms one can never be sure that lower values have not been seen only because the required conditions have not prevailled within the period for which one has data. Whilst it is almost certainly true that there is likely to always be some flux emergence (which means that there would always be some open flux and a non-zero near-Earth IMF) there is, as yet, no known physical reason that would allow one to quantify a minimum floor value. The estimate of 4.6 nT by Svalgaard and Cliver (2007a) was based on the fact that their reconstruction (SC07) did not go below this value. In fact, in the recent low solar minimum, annual values of B fell to 3.9 nT in 2009 and so Svalgaard and Cliver (2010) (SC10) revised their floor value down to 4.0 nT (which is the lowest value for calendar years), which is also shown in Figure 30. Subsequently, Cliver and Ling (2011) have generated a new estimate of a floor IMF value of B about 2.8 nT in annual means, based on more sophisticated empirical arguments, but the physical origin of any such a quantified limit remains unknown. The middle panel of Figure 30 shows the open solar flux reconstructions and the floor values have been mapped from the upper panel using the polynomial fit shown in Figure 29. One point to note is that a linear fit to the data shown in Figure 29 does not set a floor value at the intercept. The reason is that this intercept is at $F_S = 0$ and $B \approx 2\,\mathrm{nT}$. No source for the near Earth IMF, other than the coronal source flux, has ever been suggested and, hence, if $F_S = 0$ then $B = 0$ also. Hence, if a linear fit to Figure 29 is argued to be evidence for a floor value, then an explanation of where the $B \approx 2\,\mathrm{nT}$ come from, as it cannot be from the Sun. Much more realistic is that it does come from the Sun and that the relationship between open solar flux F_S and near Earth IMF B is not linear. In Section 10, the non-linear fit in Figure 29 is used to estimate the open solar flux during the Maunder minimum from cosmogenic isotope data.

All the reconstructions in Figure 30 show general variations with sunspot number, not just over the solar cycle but on centennial scales as well. The key difference between the sunspot variation and those derived for B and F_S is that sunspot activity indices return to a value close to zero every minimum (not exactly zero, there is a small long term drift in the minimum values that mirrors those in the maxima and in the 11-year running means). In contrast, both B and F_S show variability in the cycle minimum values which almost matches that in the solar maximum values. The realisation that the Sun does not return to the same baselevel state at each solar cycle minimum, even though it is (almost) clear of spots then, is an important change in our understanding of long-term solar variability. In using the two reconstructions (B and F_S), two points should be remembered. (1) The open solar flux has the advantage of being a global value that applies to the whole heliosphere whereas the IMF is a local value that applies only near the Earth (so, for example, it varies as the solar wind speed increases/decreases, making the Parker spiral unwind/tighten, respectively). (2) On the other hand, mapping from the near-Earth measurements back to the coronal source surface causes, as discussed in Section 7, its own complications and uncertainties. Hence, the IMF has the advantage of being much more straightforward observationally.

As noted above, the major differences between the reconstructions is before 1880, for when the SC10 is slightly, but consistently, higher than the other reconstructions, thereby giving less long-term trend and a higher floor value (at least over the period since 1835). Note, however, that the other reconstructions do still (just) agree within the computed uncertainty in SC10. Considerable effort is being expended deploying more datasets to try to resolve this discrepancy. However, I urge some caution here. Some of the early data are of higher quality and better long-term stability

than others, and so great care must be taken to ensure bad data is not used to corrupt good data. The author's personal view is that it may well be better to look at the present and the future to evaluate the reconstructions. As predicted as early as 2005 from the observed polar fields by Svalgaard *et al.* (2005), cycle 24 is proving to be a very weak cycle (e.g., Lockwood *et al.*, 2012) and it is instructive to look at the latest 12-month means available at the time of writing (31 March 2013). These are shown by the black dots in the three panels of Figure 30. The B and F_S values are taken from the IMF observations by the ACE spacecraft and the value of R_g is taken from the daily means of the International Sunspot Number (compiled by SIDC, Belgium), linearly regressed against R_g for the years when both are available. All the current indications are that this value is close to the maximum value (Lockwood *et al.*, 2012) which means that the current cycle (number 24) is similar in magnitude to cycle number 14 (which peaked around 1908). Hence, it is illuminating to compare the current observed values of B and F_S with the reconstructions for close to the peak of cycle 14. The best agreement is with LEA09 (in red) and the SC10 is already significantly higher at this time – a trend that continues as one goes back in time. Thus, the recent long and low minimum between solar cycles 23 and 24 (Russell *et al.*, 2010; Lockwood, 2010) and the weakness of cycle 24 thus far (Lockwood *et al.*, 2012) are likely to discriminate between the reconstructions much more effectively than the implementation of many corrections of the pre-1900 data. The evolution of cycle 24 will be monitored and updated in Section 12. The sunspot numbers seen already in cycle 24 are still considerably larger than were seen during the Dalton minimum (marked DM in the bottom panel of Figure 30) and, of course, the Maunder minimum (MM), it therefore seems highly unlikely indeed that B and F_S did not dip under any minimum values in data recorded after 1835.

9.4.2 Analysis of uncertainty

The homogeneous construction of the $IDV(1d)$ composite by Lockwood *et al.* (2013a) allows a detailed analysis of uncertainties in the reconstructions that are based on it. In evaluating these uncertainties we need to allow for errors in both the interplanetary data and in the geomagnetic index, their effect on the regression fits and the subsequent effect on the reconstructions. Lockwood *et al.* (2013b) have carried out a comprehensive evaluation of errors in the reconstruction of IMF B from $IDV(1d)$.

The largest error in the interplanetary data is associated with the fact that the geomagnetic index responds to $B\sin^4(\theta/2)$ (or some equivalent coupling function that quantifies the southward IMF component in GSM) but we are attempting to reconstruct B. As discussed in Section 5, the average of the ratio of the two tends to a constant on annual time scales, but part (c) of Figure 10 demonstrates that there is an error associated with employing this average that is of order 10%. There is also a much smaller measurement error which has been estimated from comparisons of measurements of B from different spacecraft to be of order 0.2 nT.

Figure 31 presents an analysis of the errors in $IDV(1d)$. Because IDV is compiled from over 50 stations in modern times, it is reasonable to assume that most of the differences between $IDV(1d)$ and the appropriately scaled IDV are due to errors in $IDV(1d)$, hence the distribution of the residuals of the fit of IDV onto $IDV(1d)$ gives us an uncertainty estimate in $IDV(1d)$. This distribution for the space age is shown in Figure 31 and has a standard deviation of $\sigma = 0.459\,\mathrm{nT}$.

Lockwood *et al.* (2013b) use a Monte-Carlo method to carry out a non-linear regression fit between $IDV(1d)$ and B and evaluate the uncertainties. The points shown in Figure 32 are annual means (with piecewise removal of data during datagaps and, hence, the parameters are denoted with a prime) which are fitted with a polynomial of form given in Equation (9).

$$B_p = \chi.(IDV(1d)' - \beta)^{\alpha} \tag{9}$$

In each fit, the values of χ, β, and α that yield the minimum r.m.s. difference between the

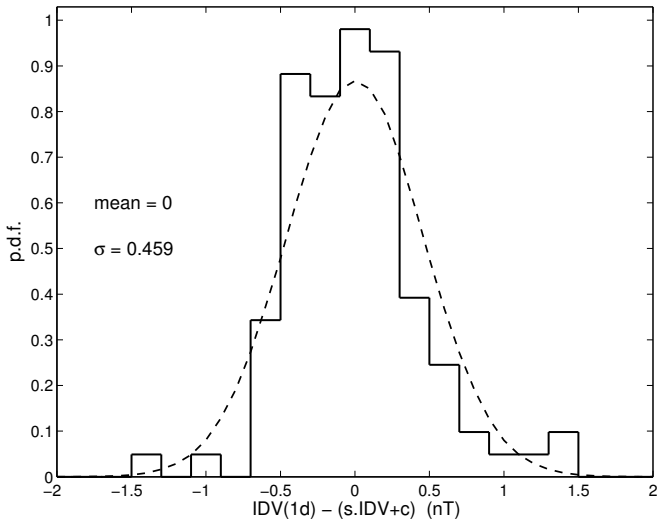

Figure 31: The distribution of fit residuals $IDV(1d) - (s.IDV + c)$ for 1966 – 2012, where s and c are the best-fit linear regression coefficients. The mean of the distribution is 0nT to within 5 decimal places. The standard deviation is 0.459 nT. The dashed line is the best-fit Gaussian distribution of the same mean and standard deviation.

observed and predicted IMF values (B' and B_p, respectively) are determined using the Nelder–Mead search method (Nelder and Mead, 1965). This fit was carried out 100 000 times, each time each point being perturbed individually by randomly-selected errors in both B' and $IDV(1d)'$, such that the errors in B' follow the normal distribution shown in part (c) of Figure 10 and the errors in $IDV(1d)'$ follow the normal distribution shown in Figure 31. An additional error, drawn at random from a normal distribution of standard deviation 0.2 nT is added to B' to allow for IMF measurement uncertainties. For the full range of potential $IDV(1d)$ values, the median, 95-percentile, and 5-percentile were evaluated from the 100 000 fits and taken to be the best fit (the blue line in Figure 32) and the 2σ uncertainty limits (which bound the grey area in Figure 32), respectively. The correlation between B' and the best-fit B_p from $IDV(1d)'$ is 0.947. The maximum possible correlation is set by the correlation between B and $B\sin^4(\theta/2)$, which is 0.957 and hence of the unexplained variation of $100(1 - 0.947^2) = 10.3\%$, $100(1 - 0.957^2) = 8.4\%$ is caused by the variation oin the IMF orientation factor.

The uncertainty band is wide at low values of $IDV(1d)$ as there are no data to constrain the fit there. The procedure does produce quasi-linear fits (for which α is close to unity), but these are rare in the ensemble and so are close to, or beyond, the 2σ level. These linear fits produce a non-zero intercept in B when $IDV(1d)$ falls to zero. This would mean that geomagnetic activity falls to zero when the annual mean B falls below about 3nT (in annual means) and there is no known reason why this would occur. In contrast, the best non-linear fits give an intercept in $IDV(1d)$ if B fell to zero: this does make sense as it means that there is a baselevel level of geomagnetic activity driven by solar wind buffeting and phenomena such as Kelvin–Helmholtz waves on the boundary, on top of which reconnection-driven effects are added.

Because it delineates the 2σ points, then 90% of the observed data points should lie within that grey band in Figure 32 if the error estimations are correct (with 5% above the band and 5% below the band). In fact, this is true for 22 out of the 30 data points (73%). However, there is an additional factor which has not yet been allowed for which is a factor in the fit to the space-age data but which would not be a factor in reconstructing the IMF from the $IDV(1d)$ composite. Although

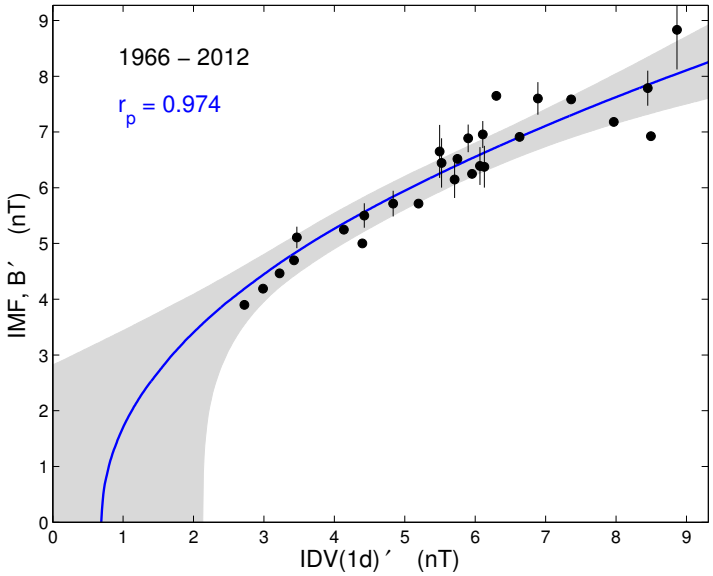

Figure 32: Scatter plot of annual means (with piecewise removal of datagaps) of $IDV(1d)'$ and IMF B', showing the results of a polynomial fit of the functional form $B = \chi.(IDV(1d)' - \beta)^{\alpha}$. The blue line is the median of 100 000 best polynomial fits and grey area defines the 2σ uncertainty band, derived using a Monte-Carlo technique, allowing for the distributions of uncertainties introduced by the IMF orientation factor $sin^4(\theta/2)$ and the experimental uncertainties in both B and $IDV(1d)$. The linear correlation coefficient of fitted and observed IMF is 0.947. The error bars on datapoints allow for the effect of the datagaps.

data gaps have been allowed for by piecewise removal of data, they still have an effect because there are (semi)annual and UT variations in the geomagnetic activity response to a given set of interplanetary conditions due to the effects of Earth's dipole tilt. If we have full data coverage, these variations are not a factor as they are averaged out in annual means. However, datagaps mean they will have an effect, depending on the UT and time-of-year at which those datagaps occur. To simulate this, the ratio of annual means of B' and B_p was evaluated for the continuous interplanetary data after 1995 but with data gaps synthetically introduced at random in such a way as to reproduce the observed distribution of gap durations in the OMNI2 dataset. Repeating this many times over allows statistical evaluation of the uncertainty in B_p caused by gaps in the IMF data, as a function of the total data coverage. Using the observed coverage, uncertainties can be assigned to annual means and these are shown by the error bars in Figure 32. Allowing for these error bars, 27 of the 30 (90%) are consistent with the grey band and this meets the 2σ design criterion.

Figure 33 shows the reconstruction and its uncertainty from this fit. The tacit assumption is that the relationship between the $IDV(1d)$ index and B found in the space age (as shown in Figure 32) applies at all other times. This is where the fact that the construction of $IDV(1d)$ is homogeneous is so important as it gives the greatest possible confidence that this is true. The best-fit reconstruction of B using the polynomial fit is the black line. The grey area surrounding this black line is the uncertainty band associated with this and is derived using the grey band in Figure 32. In addition, the uncertainties introduced into $IDV(1d)$ by the intercalibration of the stations are allowed for. This is achieved by applying the upper 2σ fit to the upper limit of $IDV(1d)$ and the lower 2σ fit to the lower limit of $IDV(1d)$. As discussed above, these confidence limits are defined to be at the 95% level. For comparison, the red line shows the result of using the linear

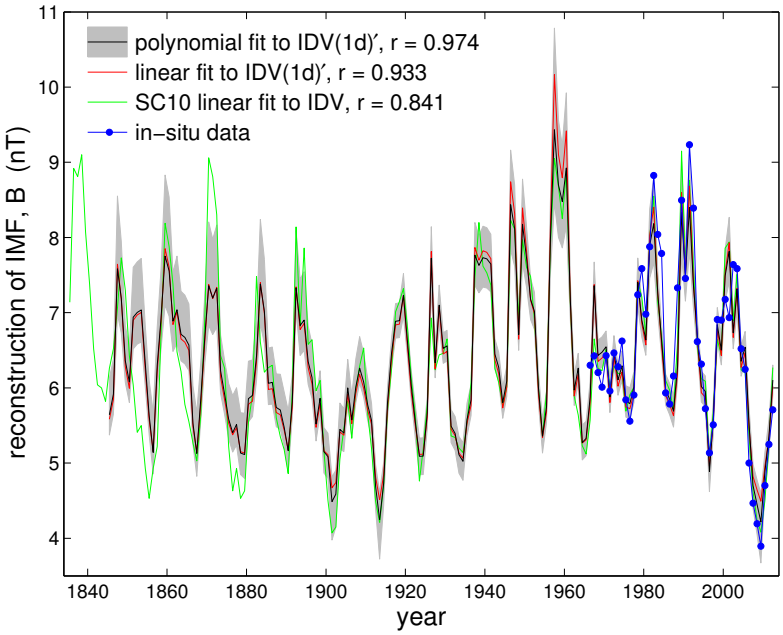

Figure 33: Reconstructions of the near-Earth IMF, B, from geomagnetic data with uncertainty analysis. The black line uses the new geomagnetic activity composite, $IDV(1d)$ (Lockwood *et al.*, 2013a) and the polynomial fit to B. The grey area surrounding this black line is the uncertainty band associated with using this polynomial fit derived using a Monte-Carlo technique (see Figure 32 and text for details) and also includes the uncertainty caused by the inter-calibration of the $IDV(1d)$ stations. The red line shows the best reconstruction using a linear fit. The green line shows the reconstruction of Svalgaard and Cliver (2010). Blue dots show the annual means of the observed IMF (from Lockwood *et al.*, 2013b).

fit: it is very similar to the results of the polynomial fit for the observed range of $IDV(1d)$. The green line shows the reconstruction of Svalgaard and Cliver (2010), including the early extension using Bartels' u index. The blue dots show the annual means of the IMF data. It can be seen that agreement between the two reconstructions is exceptionally good between 1880 and the present day (including 1901). This is despite the fact that different geomagnetic indices and different fit procedures were used by Lockwood *et al.* (2013b) and Svalgaard and Cliver (2010). Therefore, there is a real and strong consensus about the IMF reconstruction after this date. However, before 1880 there are some differences. Before 1872, the Svalgaard and Cliver (2010) reconstruction is using the Bartels' u index about which Bartels himself expressed some reservations. On the other hand, the Lockwood *et al.* (2013b) reconstruction is based on data from the Helsinki observatory (Nevanlinna, 2004) which has passed a number of self-consistency checks (Lockwood *et al.*, 2013a) and is very well correlated with corresponding data from Russian observatories operating at the same time (Nevanlinna and Häkkinen, 2010).

10 Comparison with Cosmogenic Isotopes

Galactic cosmic rays hitting Earth's atmosphere generate radionuclides by spallation (Beer *et al.*, 2012). Some of these cosmogenic isotopes are stored in terrestrial reservoirs (notably ^{10}Be in ice sheets and ^{14}C in tree trunks), into which dateable cores can be drilled. Because the flux of cosmic rays is modulated by the heliospheric field (Parker, 1965; Potgieter, 1998, 2013), the abundances of these isotopes gives unique information on the long-term variability of the Sun (O'Brien, 1979; Stuiver and Quay, 1980; O'Brien *et al.*, 1991; Beer, 2000; Muscheler *et al.*, 2007; McCracken and Beer, 2007; McCracken, 2007; Solanki *et al.*, 2004) once the effects of the secular variation in the geomagnetic field (which also shields Earth's atmosphere from cosmic rays) have been accounted for (Bhattacharyya and Mitra, 1997; Masarik and Beer, 1999). Comprehensive reviews of the methods and the underpinning science are present in Usoskin (2013) and Beer *et al.* (2012). It is useful to employ both ^{10}Be and ^{14}C because the deposition into their respective reservoirs is completely different and checking for close agreement between the inferred production rates can eliminate the possibility of signals in the record caused by changes in Earth's climate (Bard *et al.*, 1997). Gleeson and Axford (1968) showed, with some approximations, that cosmic rays behave as if they were modulated by an electric field that shields them away from the inner heliosphere. This led to the concept of the solar modulation potential ϕ which is now thought of as a parameter (in units of MV) which describes the heliospheric modification of the local interstellar spectrum (LIS) of galactic cosmic rays at the Earth (Caballero-Lopez, 2004; Usoskin *et al.*, 2005). Note that ϕ increases with increased levels of solar activity such that the fluxes of cosmic rays at Earth fall. An excellent review of long term variability of the Sun and heliosphere, as derived from cosmogenic isotopes is given in the Living Review by Usoskin (2013) and so that material will not be repeated here. However, cosmogenic isotopes do provide an independent way of testing (and extending back in time) the reconstructions presented here and so a brief comparison is worthwhile.

Lockwood (2001) noted that the open solar flux reconstruction of Lockwood *et al.* (1999a) overlapped with cosmogenic isotope records, in a way that modern data on cosmic rays from neutron monitors (Simpson, 2000) did not. A good anticorrelation was found by Lockwood (2001, 2003) on both solar cycle and centennial timescales, with the upward drift in open solar flux reflected in the downward drift in cosmogenic isotope abundances in terrestrial reservoirs, and also the drift in results from early ionisation chambers (Forbush, 1958; Neher *et al.*, 1953; McCracken and McDonald, 2001). This trend can also be detected in the cosmogenic ^{44}Ti isotope found in meteorites (Bonino *et al.*, 1995; Taricco *et al.*, 2006; Usoskin *et al.*, 2006) which is significant as it finally removes any possibility that the trend is associated with climate change influence on deposition into terrestrial reservoirs. Usoskin *et al.* (2006) use the ^{44}Ti isotope data to give strong support to models of the evolution of heliospheric fields based on sunspot number (first introduced by Solanki *et al.*, 2000), as discussed Section 11. This work showed that the well-known Hale cycle variation in cosmic ray fluxes detected using neutron monitors (with alternately peaked and then plateau-like maxima at sunspot minimum) was also well matched by the inverse of the open solar flux variation (see, in particular, Figure 2 of Rouillard and Lockwood, 2004). The anticorrelation with near-Earth IMF had been noted by Cane *et al.* (1999) and Belov (2000). Furthermore, Thomas *et al.* (2013) has shown that this feature is also present in the B and F_S reconstructions from geomagnetic activity. This raises an interesting question, which remains largely unresolved, as to the relative influences of cosmic ray drifts in the heliosphere and of the open solar flux on the modulation of cosmic rays arriving at Earth, both on decadal and centennial time scales. That the open solar flux is a factor is not a surprise as cosmic rays are scattered off irregularities in the heliospheric field and those irregularities are known to scale in amplitude with the average field value and, as shown for near-Earth space by Figure 29, that field scales with the F_S. The drift theory is very well established (e.g., Jokipii *et al.*, 1977; Jokipii, 1991; McDonald *et al.*, 1993) and has some notable successes; for example, the antiphase Hale cycle seen in electrons (Evenson, 1998)

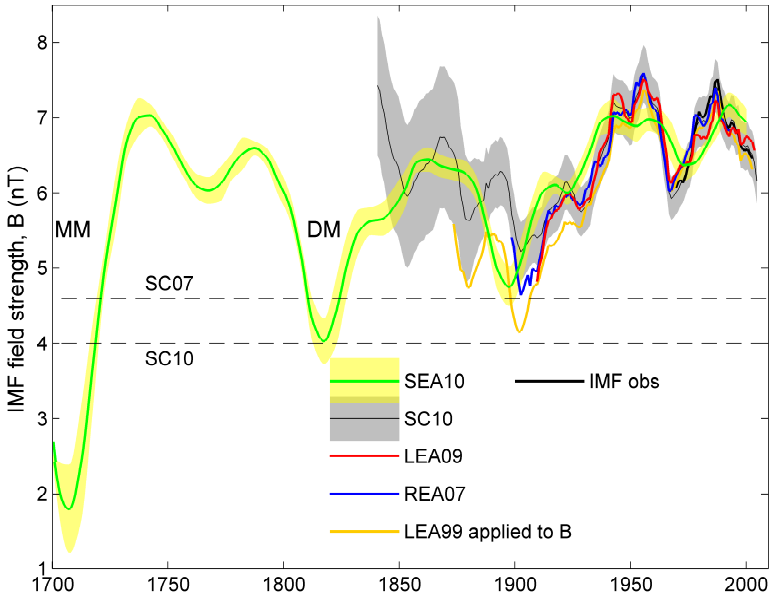

Figure 34: Long-term variations of 25-year means in the IMF B derived from ^{10}Be cosmogenic isotope data by Steinhilber *et al.* (2010) (SEA10): the green line is the best estimate and the yellow band the estimated uncertainty. In addition, 11-year running means of the reconstructions from geomagnetic activity, as presented in Figure 26, are shown using the same colour scheme as in that figure. Also shown are the SC07 and SC10 floor estimates for annual means and the Dalton and Maunder sunspot minima are labelled DM and MM. Image reproduced by permission from Lockwood and Owens (2011), copyright by AGU.

and positrons (Clem and Everson, 2002) and their latitudinal variations (Heber *et al.*, 1999). If it is assumed that these drift effects contribute to the Hale cycle but not the secular drift, their effect can be averaged out by taking means over the Hale cycle (Steinhilber *et al.*, 2008). Using ice core records of the abundance of the ^{10}Be cosmogenic isotope and a simple theory of cosmic ray shielding, Steinhilber *et al.* (2010) (SEA10) have reconstructed 25-year means of the IMF B over the last 9300 years. The results since the Maunder minimum are shown in Figure 34 and compared with 11-year running means of B from the reconstructions discussed in Section 9.1.

The general agreement between the geomagnetic and cosmogenic isotope reconstructions is extremely good although there are obvious differences and there may be some timing errors which may turn out to be attributable to dating problems with the ice cores. The agreement is very good after 1900 but less good before then. Between 1850 and 1875 the SC10 reconstruction agrees well with the average level of the SEA10 reconstruction, although showing oscillations that are not found in the SEA10 data. However, SC10 yields higher values of B in the intervals 1875 – 1905 and 1835 – 1850. The 25-year means of the SEA reconstruction of B remain above the SC10 postulated floor level for annual means, even in the Dalton minimum (DM). However, this is not true of the Maunder minimum (MM) where they fell well below it. Even in 25-year means the SEA10 B estimate fell to 1.80 ± 0.59 nT by the end of the Maunder minimum, which is still lower than the downward revision of the floor estimate to 2.8 nT by Cliver and Ling (2011). Extending the sequence over 9300 years, SEA10 find 14 grand solar minima in which the reconstructed B fell to even lower values in 25 year means. The SEA10 value of B (and its uncertainty) at the end of the Maunder minimum is marked by the white dot in Figure 29, and using the polynomial fit

shown, this yields an estimate of the signed open solar flux at the end of the Maunder minimum of $(0.48 \pm 0.29) \times 10^{14}$ Wb.

The open flux continuity model discussed in Section 11, which was derived to explain and fit the open solar flux reconstructions from geomagnetic activity data, has been used to estimate the variation of sunspot numbers from the cosmogenic data isotope data for the last millennium (Usoskin *et al.*, 2003; Solanki *et al.*, 2004). These studies found that the recent grand maximum contained unusually high sunspot numbers in the past 11 000 years, a conclusion that generated some debate (Raisbeck and Yiou, 2004; Usoskin *et al.*, 2004; Muscheler *et al.*, 2005; Solanki *et al.*, 2005). Using the composite of cosmogenic isotope data compiled by Steinhilber *et al.* (2008), Abreu *et al.* (2008) found that the recent grand solar maximum may not have been the largest in the sequence, but it was the longest in duration.

11 Models of Open Solar Flux Variation

The extension of the coronal field into the heliosphere and, hence, the modelling of the open solar flux variation, has been recently covered in another Living Review by Owens and Forsyth (2013). Therefore, as in the last section, only a brief review will be given here, to stress the extent to which the reconstructions are both feeding into the models and providing tests of them.

A number of theoretical concepts for the evolution of the heliospheric magnetic field have been proposed. Fisk (1999) argue that the Sun's open flux tends to be conserved, with "interchange reconnection" (see Crooker *et al.*, 2002) between open and closed solar fields resulting in an effective diffusion of open flux across the solar surface without, necessarily, any net change in the total open flux. In this case, the heliospheric field evolves with simple rotation of regions of positive and negative polarity separated by a single, large-scale heliospheric current sheet (Fisk and Schwadron, 2001; Jones *et al.*, 2003). That this "Fisk circulation" can conserve the open flux does not mean there are not other processes that act simultaneously to cause it to grow and decay. It has been argued that emerging midlatitude bipoles cause closed coronal loops to rise and first destroy pre-existing open flux in the polar coronal hole (remnant from the previous solar cycle) and then build up a new polar coronal hole (of the opposite polarity) and so reverse the polar field of the Sun (Babcock, 1961; Wang and Sheeley Jr, 2003b), which fits well with the migration of photospheric fields seen in magnetograph data (see the Living Review by Sheeley Jr, 2005). The evolution of the heliospheric magnetic field could also be facilitated by transient events (Low, 2001): specifically, Owens and Crooker (2006, 2007) and Owens *et al.* (2007) investigated the role of the magnetic flux contained in coronal mass ejections (CMEs) in the observed variation in flux seen by craft in the heliosphere. These different concepts are not mutually exclusive in many respects (see review by Lockwood, 2004). One complicating factor in this debate has been semantics: "open flux" in this review, and in many previous papers, is taken to be the same as coronal source flux; that is, the magnetic flux that leaves the solar atmosphere and enters the heliosphere by threading the coronal source surface at $r = 2.5\,R_{\odot}$. As discussed in this review, it is a measurable quantity because of PFSS modeling (within the assumptions of that technique) and because of the Ulysses result allows the use of in situ magnetic field data. This is quite different from another definition of open flux which requires that it has only one footpoint still (effectively) attached to the Sun (e.g., Schwadron *et al.*, 2008). Flux which appears to be in this category can sometimes be inferred for in-situ point measurements, for example, from heat flux or unidirectional strahl electron distribution functions (although scattering by heliospheric structure into other populations such as halo often makes this far from unambiguous) (Larson *et al.*, 1997; Fitzenreiter *et al.*, 1998; Owens *et al.*, 2008b). Even if this could be done reliably, there is no way to quantify the total of such flux at any one time from such in situ point measurements. This is because there is no equivalent of the Ulysses result to generalize in situ point measurements into a global quantity. Lockwood *et al.* (2009c) have reviewed how various phenomena (coronal mass ejections, interchange reconnection, disconnection reconnection) influence both these two definitions of open flux.

The long-term change in the open flux (meaning coronal source flux, F_S) deduced from geomagnetic activity has been reproduced by a number of numerical models of flux continuity and transport during the solar magnetic cycle, given the variation in photospheric emergence rate indicated by sunspot numbers (Solanki *et al.*, 2000, 2002; Schrijver *et al.*, 2002; Lean *et al.*, 2002; Mackay and Lockwood, 2002; Wang and Sheeley Jr, 2002; Wang *et al.*, 2002; Wang and Sheeley Jr, 2003b; Wang *et al.*, 2005). The key fundamental principle was established by Solanki *et al.* (2000), namely the continuity of total open solar flux:

$$\mathrm{d}F_S/\mathrm{d}t = S - L\,, \tag{10}$$

where t is time, S is the open flux emergence rate (the source term), and L is the total open flux loss term. This simple equation works without considering the precise distribution of open flux

over the Sun and how it changes, although that will undoubtedly influence the loss rate greatly and could influence the source term as well. In order to extend the modelling back to the Maunder minimum, the group sunspot number, R_g has generally been used in some form to quantify S. An initial concern was that even a model as simple as Equation (10) may have too many free variables to be meaningful. However, it should be noted that the model was "trained", and the coefficients defined, using the LEA99 open solar flux reconstruction that extended up to 1995. Therefore, although the first perihelion pass of Ulysses (between day 280 of 1994 and day 235 of 1995) was not independent data, it is significant that the model reproduced the open solar flux detected during the second perihelion pass (between day 353 of 2000 and day 301 of 2001, i.e., roughly half a solar cycle later) (Lockwood, 2003) and the third perihelion pass (almost a full solar cycle later). The model, therefore, has real predictive capability. Solanki *et al.* (2002) extended the modelling by adding more classifications of closed solar flux, each governed by its own continuity equations, to define the emergence rate S. This has been very useful in allowing centennial reconstructions of total and spectral solar irradiance which are, ultimately, constrained by the open solar flux reconstructions from geomagnetic activity.

Various forms have been used for the loss rate L. Solanki *et al.* (2000, 2002) and Vieira and Solanki (2010) used a constant fractional loss, i.e., $L = F_S/\tau$ where τ is the loss time constant. In addition, a constant absolute loss rate has been used by Connick *et al.* (2011) and a fractional loss rate that varies over the solar cycle by Owens *et al.* (2011a). In particular, working from the assumption that the source term S was set by the CME emergence rate, Owens and Lockwood (2012) showed that the loss rate needed was cyclic over the solar cycle and was very well correlated with the tilt of the heliospheric current sheet, as predicted theoretically by Sheeley Jr and Wang (2001) and Owens *et al.* (2011a), and consistent with the observations that streamer belt disconnection events, as seen in coronograph images, tend to occur where the current sheet is tilted (Wang *et al.*, 1999b,a; Sheeley Jr and Wang, 2001).

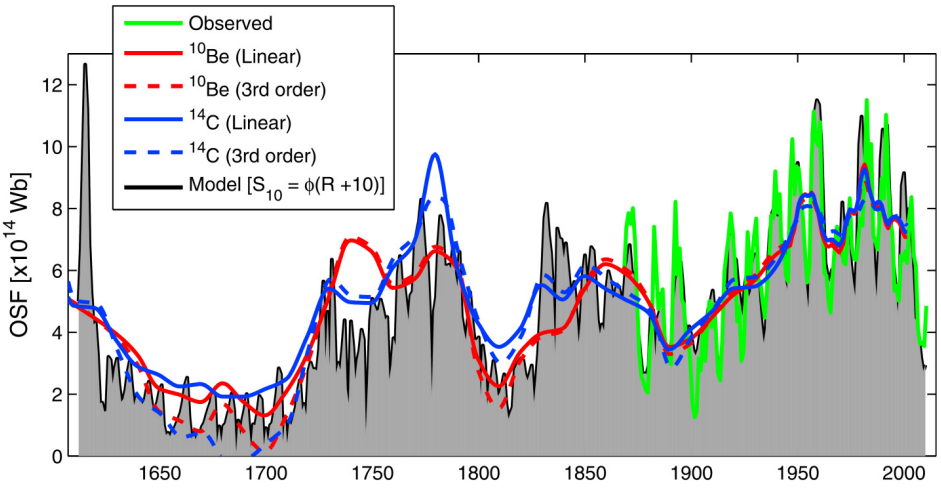

Figure 35: Long-term variation of unsigned open solar flux, $2F_S$. The grey area bounded by a black line is a model fit to the green line which is the LEA09 reconstruction. Ten year averages of cosmogenic isotope estimates of open solar flux are shown in blue (from ^{14}C) and red (from ^{10}Be), with solid and dashed lines showing linear and third-order fits of the heliospheric modulation potential variation to the open solar flux reconstruction from geomagnetic activity. Image reproduced by permission from Owens and Lockwood (2012), copyright by AGU.

Owens and Lockwood (2012) noted that during the long low minimum between cycles 23 and 24, CME flux emergence continued, and they postulated that this was a base-level emergence rate that would have continued at all times, including during the Maunder minimum. Using this they derived the modelled variation of open solar flux shown by the grey area Figure 35. Note that the plot shows the unsigned open flux and so is $2F_S$. The modelled signed open solar flux at the end of Maunder minimum oscillated around a mean of about $F_S = 0.7 \times 10^{14}$ Wb. This is a bit larger than, but still consistent within uncertainties, with the estimate from the SEA10 reconstruction and Figure 29 of $(0.48 \pm 0.29) \times 10^{14}$ Wb. The geomagnetic reconstruction has also been extended back by cross-correlating decadal means with the corresponding decadal means of the heliospheric modulation potentials derived from cosmogenic isotope abundances. This has here been done for both linear and third-order fits (solid and dashed lines, respectively) and for both ^{10}Be and ^{14}C cosmogenic isotope records (red and blue, respectively). The model is close to all these empirical extrapolations. Using the polynomial fit shown in Figure 29, the average annual mean $F_S = 0.7 \times 10^{14}$ Wb for the end of the Maunder minimum modelled by Owens and Lockwood (2012) yields an IMF of $B = 2.5$ nT. This is quite similar to the floor estimate of Cliver and Ling (2011) of 2.8 nT. Hence, the baselevel CME emergence rate postulated by Owens and Lockwood (2012) comes close to offering a potential explanation of at least the smaller floor estimates. As shown by Figure 35 this postulate matches cosmogenic isotope data quite well; however, it remains only a postulate. Note that, in addition, that the cosmogenic isotopes tell us that the Maunder minimum is not the lowest level of solar activity reached in the last 9300 years, and so it remains possible that this minimum still does not set a genuine floor limit to the IMF.

Interestingly, the modelled open solar flux shows cyclic variations during the Maunder minimum. The long time constants of exchange of carbon with the two great terrestrial reservoirs (the biomass and the oceans) means that solar cycle variations cannot be seen in ^{14}C data, but the same is not true for ^{10}Be (Beer *et al.*, 1990). One puzzling observation had been that ^{10}Be continued to show decadal-scale oscillations during the Maunder minimum when no evidence for a magnetic cycle can be found in sunspot data (Beer *et al.*, 1998). Even more puzzling was that, whereas the ^{10}Be abundances at other times are, as expected, in antiphase with sunspots numbers, at the start and end of the Maunder minimum, the ^{10}Be oscillations are in phase with the (small) sunspot cycles (Usoskin *et al.*, 2001). Owens *et al.* (2012) show how the modelling presented in Figure 35 provides the first viable explanation of these two puzzles: at high solar activity levels the solar cycle variation in open solar flux is dominated by the large sunspot cycle variations in the source rate S. On the other hand, at low activity, S becomes small and relatively constant and the open solar flux variations are dominated by the cyclic variations in the loss rate L with the current sheet tilt.

12 The Future

> *"It is not important to predict the future, but it is important to be prepared for it"*
> – Pericles, Athenian orator, statesman and general (c. 495 – 429 BC)

I have some difficulties with this quote as I don't believe you can prepare for something if you do not know what it is. Better is:

> *"It is not important to know the future, but to shape it"*
> – Antoine de Saint Exupéry, French writer and aviator (1900 – 1944)

which I can see can be valid in many areas of life such as sport, warfare, economics, and politics. However, it is invalid in solar-terrestrial physics. What undoubtedly does apply is:

> *"Prediction is very hard, especially when it is about the future"*
> – Niels Bohr, Danish Physicist (1885 – 1962)

But perhaps the wisest of all is a variant on Bohr's quote:

> *"Never make predictions, especially about the future"*
> – Lawrence ('Yogi') Berra, American Baseball Player, coach and author, (1925 –)

Ignoring the obvious wisdom of Berra's advice, this last section uses the knowledge outlined in the previous sections to look into the probable future of solar activity.

A great many papers have looked at predicting sunspot numbers, particularly those expected at the start of a new cycle, and the methods employed have been presented in the Living Review by Petrovay (2010). Because of the difficulty in making such sunspot number predictions for just the cycle ahead, the degree to which there is some predictability in several solar activity indices on longer timescales has not been exploited. Lockwood *et al.* (2011b) have used the predictability measure based on autocorrelation functions devised by Hong and Billings (1999) to show that although the predictability of sunspot numbers is indeed relatively low, it is greater over longer lags for the open solar flux and the solar modulation potential ϕ. Lockwood *et al.* (2011b) find that this predictability is great enough over sufficiently long lags to allow forecasting of the onset of a grand solar minimum such as the Maunder minimum.

The SEA10 reconstruction of near-Earth heliospheric field discussed in Section 10 was derived from the homogenised composite of the heliospheric modulation parameter ϕ from ^{10}Be abundance sequences in various ice cores compiled by Steinhilber *et al.* (2008). Figure 36 shows the full composite in 25-year means ϕ_{25}, which covers 9300 years.

The present value of ϕ_{25} is near 600 MV, which Figure 36 reveals to be high compared to the average value for this interval. The $\phi_{25} = 600$ MV level is shown by the horizontal orange line and intervals when ϕ_{25} exceeded this are shaded red. Lockwood (2010) and Barnard *et al.* (2011) use this as a threshold value to define a "Grand Solar Maxima" (GSM) as this definition means the GSM that has persisted through the space age has recently come to an end. It can be seen that such GSMs are relatively rare, indeed there are just 24 of these prior to the recent one if we adopt this definition (an average repeat period of about 390 years, but there is a large spread about this mean value as their occurrence is far from regular and these GSMs are notably more common in the second half of this composite). The times when they end (when ϕ_{25} falls back below the 600 MV level) are shown by the vertical lines. The Maunder minimum (MM) is also marked and the lowest ϕ_{25} within it $\phi_{\rm MM}$ is marked by the horizontal green line. It can be seen that the Maunder minimum is not the lowest in the sequence, and although they are more evenly spread than the GSMs, the minima are also more common in the second half of the interval. There

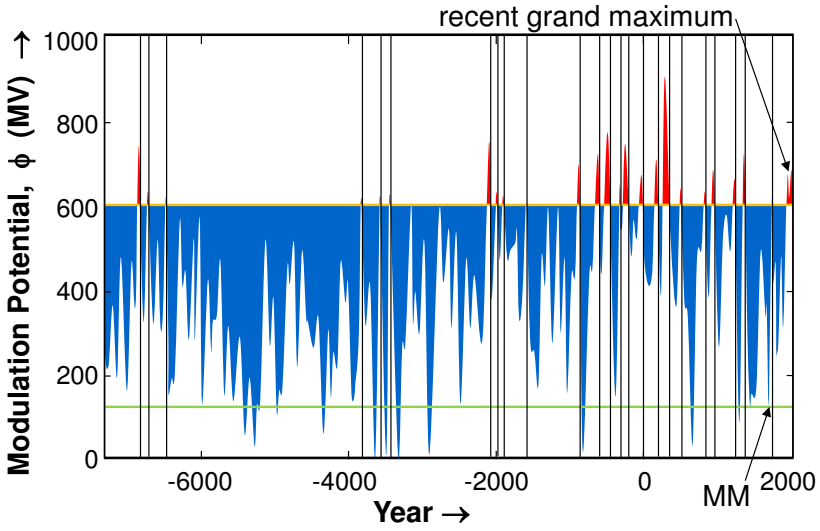

Figure 36: The composite record of 25-year means of the solar modulation potential, ϕ_{25} derived from ^{10}Be cosmogenic isotope abundances by Steinhilber *et al.* (2008). The red areas show Grand Solar Maxima (GSM) defined here as when ϕ_{25} exceeds the 600 MV level (the horizontal orange line) the blue areas are where ϕ_{25} is less than this value. The vertical lines show the end times of these GSM, t_o. The Maunder minimum is marked MM and the minimum value of ϕ_{25} during it, $\phi_{\rm MM}$, is marked by the horizontal green line.

are 12 grand minima that are at least as deep (in terms of the minimum ϕ_{25} within them) as the Maunder minimum (average repeat period 780 years) and 30 that are at least as deep as the Dalton Minimum (average repeat period 310 years). Abreu *et al.* (2008) noted that at the end of the data composite of 25-year averages (the last data point being in 1994), the Sun was still within the recent GSM but that this maximum had lasted longer than other in the past 9300 years and ϕ_{25} was currently declining, consistent with the decline commencing in 1985 noted by Lockwood (2003) and discussed in the context of global climate change by Lockwood and Fröhlich (2007). Abreu *et al.* (2008) concluded the GSM was likely to come to an end in the near future, a conclusion supported from the reconstructed and directly-observed open solar flux and IMF by Lockwood *et al.* (2009d). The discussion below shows that recent neutron monitor data reveal the recent GSM actually did end in 2001 in the 25-year means for the adopted definition.

As yet we have no predictive models of the solar dynamo that can model this time series (see the Living Review by Charbonneau, 2010) and so the only ways to use these data to predict the future remain almost exclusively empirical. Two methods have been deployed. Steinhilber and Beer (2013) have recently used two types of spectral techniques whereas Lockwood (2010) and Barnard *et al.* (2011) have made "analogue forecasts". The results of these three methods are remarkably similar and lead to the same general conclusions. This section exemplifies those conclusions using analogue forecasts, which are based on studying how ϕ_{25} has behaved in the past following a situation analogous to the present day. The situation is defined to be analogous if ϕ_{25} falls to 600 MV having previously exceeded this value: further restrictions could be applied but because the recent ϕ_{25} have been so high, even this only gives 24 previous analogues in 9300 years. The times $t = t_o$ when ϕ_{25} falls back down to 600 MV (the vertical lines in Figure 36) have been "composited" (also called a "superposed-epoch" or "Chree" analysis) such that they are all at time

zero in Figure 37, and the black lines show the variations around this time for the 24 previous GSMs.

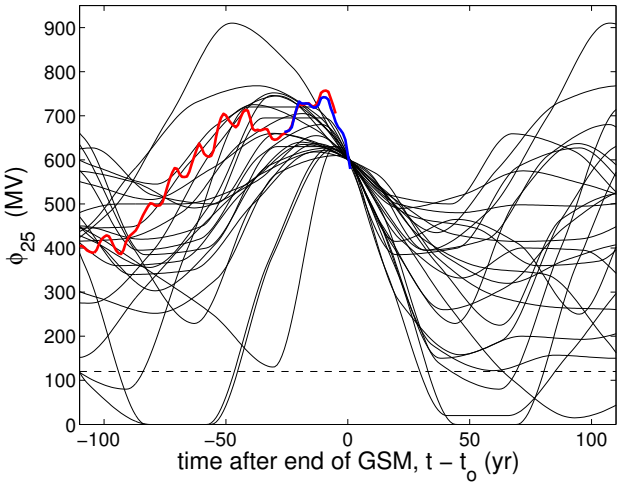

Figure 37: Update of the superposed epoch plot by Barnard *et al.* (2011). The black lines are the 25-year means of the heliospheric modulation potential ϕ_{25} derived from ice-core ^{10}Be abundances by Steinhilber *et al.* (2008) and here interpolated to annual values using splines interpolation: these are composited around the 24 times $t = t_o$ in the 9300-year record that ϕ_{25} fell below the 600 MV level. The horizontal dashed line in the Maunder minimum value, $\phi_{\rm MM}$. The blue line shows 22-year running means of ϕ derived from monthly-mean Oulu neutron monitor data with $t_o = 2001.4$ (when ϕ_{22} fell to 600 MV). The red line shows 22-year running means of the reconstruction of annual ϕ values, as derived by Usoskin *et al.* (2002) for the same t_o.

Also shown in Figure 37 are 22-year running means of the annual heliospheric modulation potential reconstruction for the past century, compiled by Usoskin *et al.* (2002) from modern neutron monitor data, ionisation chamber records and cosmogenic isotope records (red line) and of the modulation potential from the Oulu neutron monitor counts, from a third-order polynomial fit to the red line for the interval when they overlap. The horizontal dashed line is the value of $\phi_{\rm MM}$ shown in Figure 36. The blue line confirms that ϕ_{22} has recently dropped below the 600 MV level and so, by this definition, the recent GSM has come to an end. In addition Figure 37 also implies that the recent decrease in solar activity (meaning the long, low minimum between cycles 23 and 24 and the weak cycle 24 thus far) has given a decline more rapid to exit a GSM than any seen in the past 9300 years.

Lockwood (2010) noted that in two cases of the 24, ϕ_{25} fell below $\phi_{\rm MM}$ within 40 years ($t - t_o < 40$ yr) and so concluded there was a $2/24 \approx 8\%$ chance of a Maunder-like minimum in the next 40 years. We can make estimates of the probability of any level being exceeded ($P[\geq\phi_{25}] = 1 - P[<\phi_{25}]$) into the future by counting the fraction of the composited ϕ_{25} values that exceed a certain value at each epoch time ($t - t_o$). By interpolation this can be used to predict the evolution of ϕ_{25} at a given level of probability. Lockwood *et al.* (2011a) and Barnard *et al.* (2011) used empirical regressions and theoretical relations between ϕ_{25} and 25-year means of other parameters (either observed such as *aa* and R or reconstructed such as B) and then, for each, evaluated the fractional deviation of annual mean values from the 25-year means as a function of the solar cycle phase, ϵ. Hence, for a given ϕ_{25} and solar cycle phase ϵ, an annual value could be computed. Using the ϕ_{25} at a certain probability and assuming all future cycles are 11 years in duration (to prescribe ϵ), annual values of ϕ at a given probability level could be predicted into the future. The same procedure was applied to the predicted 25-year means of other parameters. The results for the

various probability levels are shown by the coloured lines in Figure 38 which also gives the past variations which are either observed (in black) or reconstructed (in mauve).

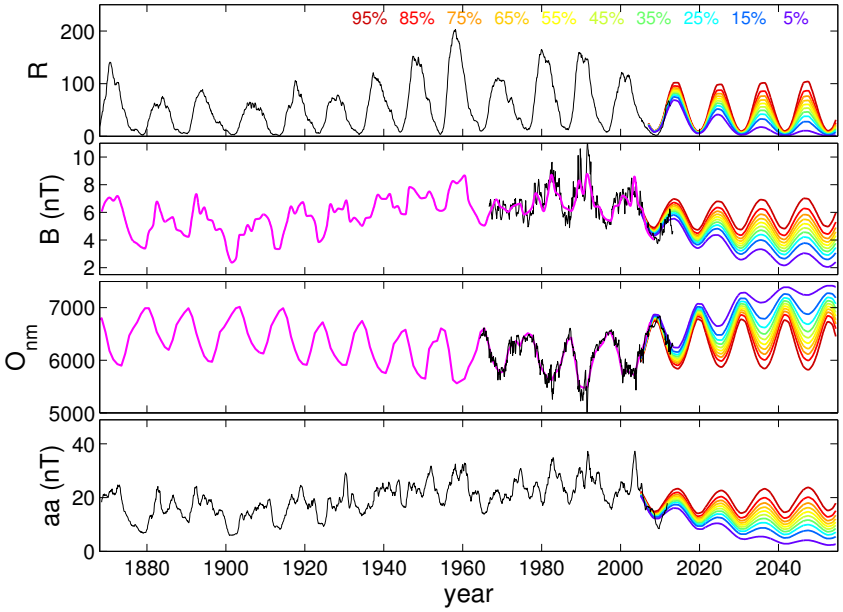

Figure 38: Observed past and predicted future variations of (from top to bottom): sunspot number, R; interplanetary magnetic field strength at Earth, B; cosmic ray counts by the Oulu neutron monitor, $O_{\rm nm}$; and the *aa* geomagnetic index. The black lines are monthly averages of observations. The mauve line in the second panel is the LEA09 reconstruction of annual means of B from geomagnetic data by Lockwood *et al.* (2009d), and that in the third panel from the reconstruction of ϕ by Usoskin *et al.* (2002). The red-to-blue lines show predicted variations of annual means at various probabilities, made from the 9300-year cosmogenic isotope composite of Steinhilber *et al.* (2008) using the procedure developed by Lockwood *et al.* (2011a) and Barnard *et al.* (2011). In the top panel the blue to red lines show the values of R which have a probability $P[< R] = [0.05 : 0.1 : 0.95]$ that R will be lower than the value shown. Corresponding predictions are given in the other panels. (Note that in the third panel the probability of $P[> O_{\rm nm}]$ is shown as $O_{\rm nm}$ rises as solar activity falls). Image adapted from Lockwood *et al.* (2012).

The plot shows that there is only a 5% probability that solar activity cycles will remain as large as or exceed recent cycles but that, at the other extreme, there is a 5% probability that they will fall to Maunder minimum levels within just under 40 years. The most likely scenario is between the yellow and green lines which places the next grand minimum some time after 2060. An interesting question becomes how is cycle 24 evolving in relation to these predictions? To answer that question, one first has to establish where in the cycle 24 we currently are.

This is here done using the method described by Owens *et al.* (2011b) who noted from the Greenwich/USAF sunspot data (Hathaway, 2010) that the variation of sunspot latitude with solar cycle phase was very similar, independent of the amplitude of the cycle. For most past cycles the mean latitude of the spots has been roughly equal in the northern and southern solar hemispheres, but a complication is that cycle 24 is proving exceptional in that the southern hemisphere is lagging considerably behind the northern (Lockwood *et al.*, 2012). As a result, the conclusion from the northern hemisphere mean sunspot latitude is that cycle 24 has passed its peak but for the southern hemisphere is that it is imminent (as of 20 April 2013).

Figure 39 takes an independent look at the evolution of cycle 24 by analysing the solar polar magnetic fields. The timing of the polar field reversal, relative to sunspot maximum, was first

observed during solar cycle 19 (SC19) by Babcock (1959) using data from the Hale Solar Laboratory (HSL) magnetograph. He noted that the average field emerging from the south solar pole reversed polarity between March and July 1957 and that in the north pole reversed in November 1958. The 12-month running mean of monthly sunspot number peaked in March 1958, midway between these two reversals. Figure 39 employs the continuous data on the solar polar field available from the Wilcox Solar Observatory (WSO). As noted by Babcock during SC19, the two poles do not reverse at exactly the same date, and the raw data are also complicated by a strong annual periodicity introduced by the annual variation in Earth's heliographic latitude. Because of these two effects, the average polar field reversals are most readily seen by taking the difference between the north and south fields, $(B_N - B_S)$. In order to give the variations of this difference the same appearance in each cycle, thereby allowing easy comparisons, the upper panel of Figure 39 shows $(B_N - B_S)$ multiplied by p, where $p = +1$ for odd-numbered cycles and $p = -1$ for even ones: the variation of $p(B_N - B_S)$ with solar cycle phase, ϵ (determined using the average of the absolute values of the northern and southern mean sunspot latitudes), is plotted in the top panel for the WSO measurements, which are made every 10 days. The area shaded gray is between the earliest (lowest ϵ) reversal which was seen during cycle 23 (green line) and the latest possible reversal date which was the brief return to $p(B_N - B_S) = 0$ during cycle 22 (blue line). (However, notice that the best estimate of the reversal for cycle 22 was at considerably lower ϵ). The lower panel shows $-pB_{fN}$ and pB_{fS} where B_{fN} and B_{fS} are the northern and southern polar field variations after they have been passed through a 20 nHz low-pass filter to smooth them and remove the annual variation. The vertical lines give the phases of the corresponding cycle peaks in 12-point running means of monthly sunspot numbers. Red, blue, and green are used to denote cycles SC21, SC22, and SC23 and black is for SC24. The Figure shows that the polar fields during SC24 thus far have been weaker than they were in the corresponding phase of the previous three cycles. It is noticeable that for the odd-numbered cycles the reversal of the poles is within roughly a month of the smoothed sunspot number peak. However, for the even-numbered cycles the polarity reversal took place considerably after the cycle peak.

Note that the predictions made by Steinhilber and Beer (2013), Lockwood (2010) and Barnard *et al.* (2011) are probabilistic rather than deterministic (or categorical) in nature. Many areas of geophysics, including weather and flood forecasting (Krzysztofowicz, 1998; Bartholmes *et al.*, 2009), have concluded that probabilistic forecasting is more powerful in decision-making, that they are usually scientifically more honest, and that for applications they enable risk-based assessments. Tests of forecast skill have been developed but probabilistic forecasts are not yet in widespread use in solar, heliospheric, and solar-terrestrial science which has remained more deterministic in approach. Their increasing use would be a natural part of the development of solar-terrestrial physics into applications-based "space weather".

Figure 40, like Figure 39, will be updated as the cycle progresses. Panels (a), (c), and (d) show monthly means (in grey) and 12-month running means (in black) of the international sunspot number R, the Oulu neutron monitor cosmic ray counts O_{nm}, and the observed IMF B, respectively. These are compared to the predictions shown in Figure 38, presented using the same colour scheme to give the probability of the parameter being lower than the value shown. Panel (b) shows the evolution of the mean sunspot latitude in the northern (in blue) and southern (in red) hemispheres. The current date is shown by the vertical black dashed line (after which the dashed lines show linear extrapolations based on the prior data for cycle 24). The circles on each line mark the latitudes when peak sunspot area in that hemisphere was observed during previous cycles.

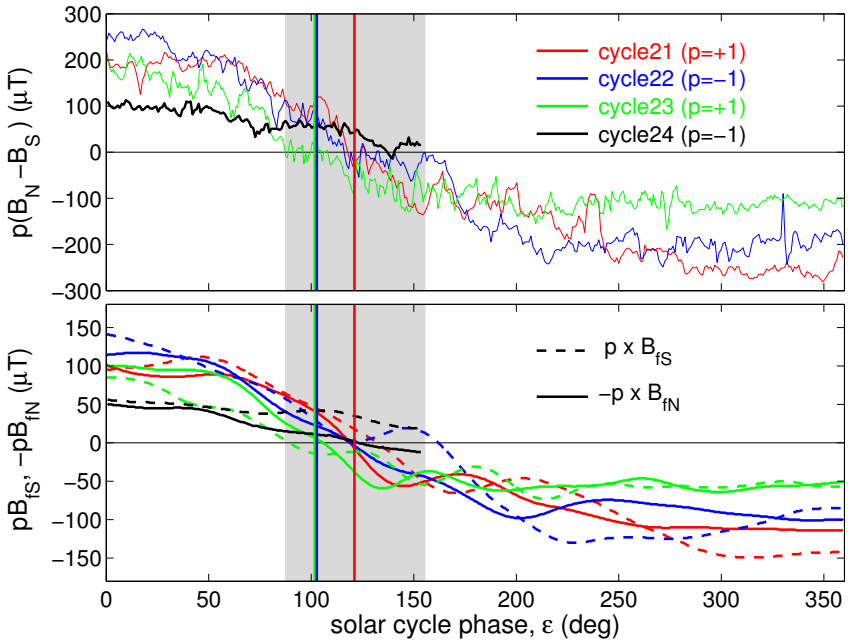

Figure 39: Solar polar fields observed by the magnetograph at Wilcox Solar Observatory (WSO) during solar cycles 21 – 24. The top panel shows the difference between the two polar fields, $p(B_{\rm N} - B_{\rm S})$ as a function solar cycle phase, ϵ, as determined from mean sunspot latitudes using the method described by Owens *et al.* (2011b), where $B_{\rm N}$ and $B_{\rm S}$ are the average fields seen over the north and south solar poles, respectively, and $p = +1$ for odd-numbered cycles and $p = -1$ for even ones. The reversals all occur within the grey band and the phases of the peak sunspot number in 12-month running means are given by the vertical lines. The lower panel shows $-pB_{\rm fN}$ (solid lines) and $pB_{\rm fS}$ (dashed lines) as a function of ϵ where $B_{\rm fN}$ and $B_{\rm fS}$ are the $B_{\rm N}$ and $B_{\rm S}$ data that have been passed through a 20 nHz low-pass filter. In both panels, red, blue, green and black denotes solar cycles SC21, SC22, SC23, and SC24, respectively. This plot will be updated regularly as the cycle progresses: this version was generated on 20 April 2013.

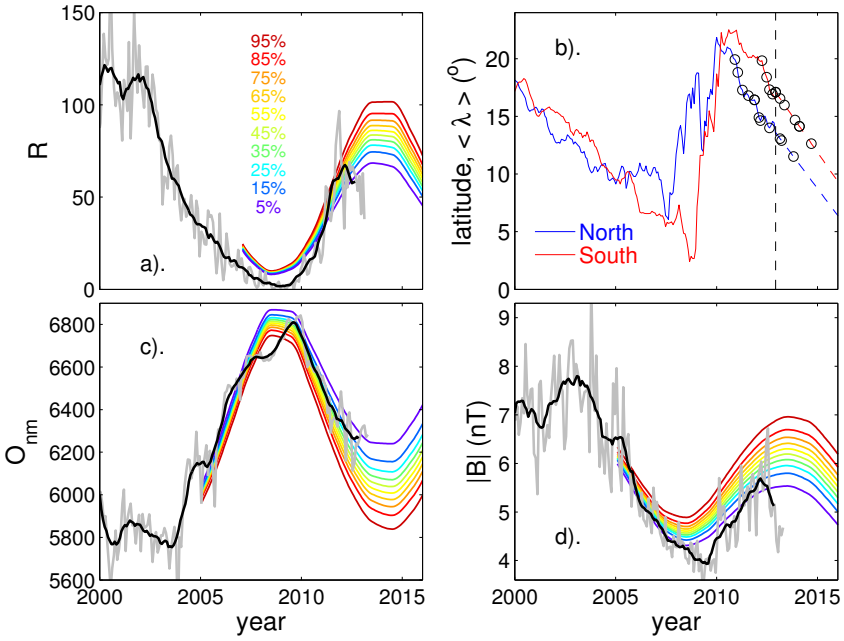

Figure 40: The evolution of solar cycle 24 thus far. Observed monthly means (in grey) and 12-point running means of monthly data (in black) of (a) sunspot number, R; (c) Oulu neutron monitor counts, $O_{\rm nm}$ and (d) the observed near-Earth IMF field strength, B. In each of these plots the coloured lines are the predicted levels at various probabilities P between 95% (in red) and 5% (in mauve), as derived by Barnard *et al.* (2011). The value shown is that which has a $(100 - P)$% chance of being exceeded in the case of R and B and a P% chance of being exceeded in the case of the comic ray counts. Panel (b) shows the monthly mean latitudes of sunspots groups $\langle\lambda\rangle$. The red line is for the southern solar hemisphere, the blue for the north and solid line is observed whereas the dashed is a linear extrapolation of the cycle 24 behaviour into the future. The circles show the mean latitude for that hemisphere at which peak sunspot number was seen in cycles 12 – 23: the number of open circles to the right of the dashed line is the number of those cycles that had not yet reached their peak sunspot number in that hemisphere for the latest shown mean latitude of spots in that hemisphere. This plot will be updated regularly as the cycle progresses: this version was generated on 20 April 2013.

12.1 Solar cycle 24 update 1: 20 April 2013

At present, all the data shown in Figure 40 (sunspot number, R, Oulu neutron monitor cosmic ray flux, $O_{\rm nm}$, and near-Earth IMF B are following the predicted blue lines. These are for low (5 – 10%) $P[< R]$, $P[> O_{\rm nm}]$, and $P[< B]$ values. Thus, the decline in solar activity is very much at the more rapid end of the range predicted from the analogue forecasts, which is consistent with an earlier onset of the next grand minimum. However, it must be stressed there is no dynamo science in these analogue forecasts and without a full understanding of the physics of the long-term changes described in this review, we can have no confidence that an observed trend will continue.

All the indicators are that cycle 24 is close to, or has past, its peak. The northern polar field has flipped and the northern hemisphere spots have migrated equatorward to a latitude below that for sunspot maximum in all but 3 of the 12 cycles for which we have data on spot latitudes. In the southern hemisphere the polar field flip appears imminent, but has yet to occur and the spots have migrated equatorward to a latitude below that for sunspot maximum in 6 of the 12 cycles. Thus, from average sunspot latitudes, the northern solar hemisphere indicates a probability of 75% that the sunspot maximum of cycle 24 has already occurred at this date and the southern hemisphere gives a probability of 50%.

12.2 Solar Cycle 24 Update 2: 1 August 2013

Figure 41 is an updated version of Figure 39 and the upper panel shows a significant development in that the polarity the difference between the northern and southern solar polar fields has reversed. If this is the final reversal (it did flip briefly earlier in the cycle) it means that is at a slightly greater phase of the cycle than we have seen before, but is only slightly later than during cycle 22. The filtered data in the bottom panel (that are effectively extrapolations for the most recent data) indicate that although the northern polar field has flipped, the southern has still yet to do so.

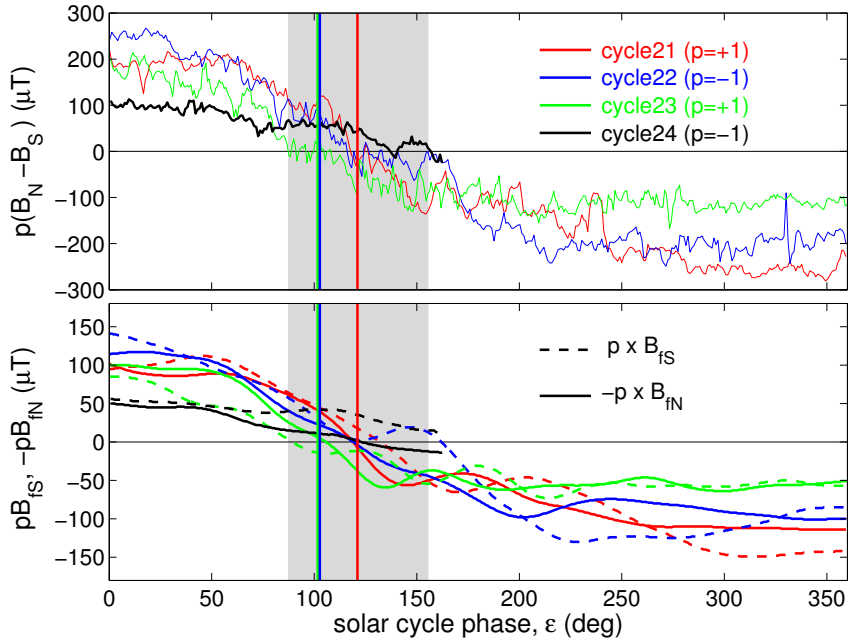

Figure 41: Update of the solar polar field plot given in Figure 39. Plot updated on 1 August 2013.

The corresponding update to Figure 40 is shown in Figure 42. Part (a) shows that in the period since the previous update, the monthly sunspot number R has shown an increase but the 12-month running means are still considerably below the 95-percentile prediction (blue line). Similarly, part (d) shows that the IMF B also remains well below this line in 12-month running means. Part (c) shows that there have been some Forbush decreases that have lowered the monthly mean cosmic ray count (mirroring the rise in R) but the 12-month running mean remains between the 90 and 95 percentiles. The average sunspot latitudes have decreased such that for the northern hemisphere only 1 of the 12 previous cycles has a lower average latitude at cycle maximum and the corresponding number for the southern hemisphere has fallen to 5. Thus, the average sunspot latitudes in the northern and southern hemispheres give probabilities of 92% and 58% that solar maximum has already been reached by this date.

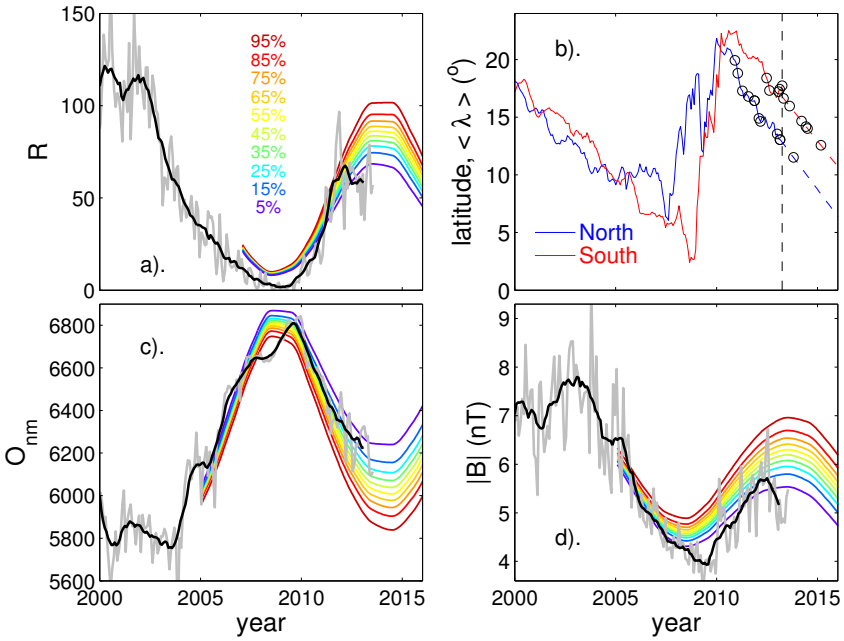

Figure 42: Update of Figure 40 summarising the evolution of solar cycle 24 thus far. Plot updated on 1 August 2013.

13 Concluding Remarks

The purpose of this review is to detail the development of reconstructions of solar and heliospheric magnetic fields from geomagnetic activity data and to make some initial comparisons with the results from cosmogenic isotopes and models based on sunspot number observations. One important aspect of this work on centennial-scale solar variability is that it fills the timescale gap between decadal-scale variations (almost 5 full cycles are now covered by in-situ space measurements) and millennia (covered by cosmogenic isotope data) and so allows modern space-age understanding to be applied to the cosmogenic isotope data in a more precise, insightful and quantitative manner.

A fundamental insight that has accrued is that although the sunspot number returns to an (almost) constant baselevel state at every solar minimum, this disguises the fact that the Sun does not return to the same state at each sunspot minimum: there are long-term variations on coronal and heliospheric fields which mean that one solar minimum is not the same as the next. The cycle-to-cycle variation of the Sun at solar minimum is the basis of the "precursor" method for predicting the peak of the sunspot cycle (Schatten *et al.*, 1978) and using the polar fields Svalgaard and Cliver (2005) made predictions for cycle 24 which are proving exceptionally accurate. This long-term change influences many space-weather phenomena, such as the galactic cosmic ray flux reaching Earth, ("gradual") solar energetic particles generated ahead of solar coronal mass ejections, and solar-wind interactions that drive solar-terrestrial activity giving phenomena such as Geomagnetically Induced Currents in power grids. Hence, we need to allow for "space climate change" as well as space weather.

Another important realisation has been that information has been lost by grouping the variations of various observation series and indices into a general, catch-all "solar activity" classification. Specifically, although the variations of the various geomagnetic activity indices have a great many similarities (for example, all reflect the decadal-scale sunspot cycle), they are measures of different parts of the currents systems in near-Earth space and they have different dependencies on solar wind parameters. These differences can be exploited to derive new information. The example discussed most here is how combinations of range and interdiurnal variation geomagnetic indices allow us to separate the effects of solar wind speed and the interplanetary magnetic field and so reconstruct both back in time.

The review has covered many pitfalls of the reconstructions from knowing the full provenance of historic data, to limitations of statistical methods. Nevertheless considerable consensus now exists between the various reconstructions for between about 1880 and the present day. Thus, using geomagnetic activity allows us to extend the 50-year sequence of in-situ spacecraft data a further 80 years back in time with great confidence. Before 1880 there are increasing differences. There were fewer stations and they were frequently in increasingly noisy environments and eventually had to be moved as cities expanded around them. Equipment was less accurate and more prone to calibration drifts. Thus, although there is, in theory, information available for a further 75 years back into the past there is necessarily greater uncertainty. This is not to say that the full 185-year record cannot be recovered, but the earliest data will always have greater associated uncertainties.

The most important realisation that these reconstructions have allowed, when combined with cosmogenic isotope data, is that the modern space age has been an unusually active period for the Sun, compared to most of the last millennium. Just how unusual is a matter of on-going debate, but even this is now converging to a consensus view that it has been, until recent years at least, very unusual indeed. That being the case, we should not be surprised that average solar activity levels are now declining and that cycle 24 is weak compared to the others in the space age. The cosmogenic isotope data tell us that this decline is more likely to continue than not. This has great implications for solar physics, solar-terrestrial science and space weather. Some effects of a continuation of the current decline are well known, for example, the cosmic ray flux incident on Earth will rise. However, others are not. It is possible that although large space weather

"events" may become fewer in number, the largest could become more severe in their terrestrial effects because the CME is ejected into a lower-field heliosphere making the Alfvén Mach number of the event greater and potentially reducing Sun-to-Earth transit times. The long, low minimum between cycles 23 and 24 was only "exceptional" in the context of the space age and may give pointers to other changes that we should now expect. The author's personal view is that this offers great scientific possibilities and the modern observation techniques applied to a quieter Sun will teach us much more than a continuation of the high activity levels seen during cycles 21, 22, and 23. All the evidence is that cycle 24 has just passed its peak, and that peak is a weak one, the development of solar activity into the next minimum will be very interesting to monitor.

14 Acknowledgments

The author thanks a great many scientists who have contributed to this text through their insights, data, discussions, and reprints. He is also exceedingly grateful to all his collaborators and, in particular, his colleagues Luke Barnard, Mat Owens, Chris Davis, Giles Harrison, and Clare Watt.

References

Abreu, J.A., Beer, J., Steinhilber, F., Tobias, S.M. and Weiss, N.O., 2008, "For how long will the current grand maximum of solar activity persist?", *Geophys. Res. Lett.*, **35**, L20109. [DOI], [ADS] (Cited on pages 59 and 64.)

AGU [Council], 2009, "The Importance of Long-term Preservation and Accessibility of Geophysical Data", *Eos Trans. AGU*, **90**(35), 301. [DOI] (Cited on page 21.)

Altschuler, M.A. and Newkirk, G., 1969, "Magnetic fields and the structure of the solar corona", *Solar Phys.*, **9**, 131–149. [DOI], [ADS] (Cited on page 32.)

Arge, C.N. and Pizzo, V.J., 2000, "Improvement in the prediction of solar wind conditions using near-real time solar magnetic field updates", *J. Geophys. Res.*, **105**, 10,465–10,479. [DOI], [ADS] (Cited on page 32.)

Arge, C.N., Odstrčil, D., Pizzo, V.J. and Mayer, L.R., 2002, "Improved Method for Specifying Solar Wind Speed Near the Sun", in *Solar Wind Ten*, Proceedings of the Tenth International Solar Wind Conference, Pisa, Italy, 17 – 21 June 2002, (Eds.) Velli, M., Malara, F., Bruno, R., AIP Conference Proceedings, 679, pp. 190–193, Amercian Institute of Physics, Melville, NY. [DOI], [ADS] (Cited on page 33.)

Arnoldy, R.L., 1971, "Signature in the interplanetary medium for substorms", *J. Geophys. Res.*, **76**, 5189–5201. [DOI], [ADS] (Cited on page 13.)

Babcock, H.D., 1959, "The Sun's Polar Magnetic Field", *Astrophys. J.*, **130**, 364–365. [DOI], [ADS] (Cited on page 67.)

Babcock, H.D., 1961, "The Topology of the Sun's Magnetic Field and the 22-year Cycle", *Astrophys. J.*, **133**, 572–586. [ADS] (Cited on page 60.)

Baker, D.N., Hones, E.W., Payne, J.B. and Feldman, W.C., 1981, "A high time resolution study of interplanetary parameter correlations with AE", *Geophys. Res. Lett.*, **8**, 179–182. [DOI], [ADS] (Cited on page 13.)

Balogh, A., Smith, E.J., Tsurutani, B.T., Southwood, D.J., Forsyth, F.J and Horbury, T.S., 1995, "The heliospheric field over the south polar region of the Sun", *Science*, **286**, 1007–1010. [DOI], [ADS] (Cited on page 33.)

Bard, E., Raisbeck, G.M., Yiou, F. and Jouzel, J., 1997, "Solar modulation of cosmogenic nuclide production over the last millenium: comparison between ^{14}C and ^{10}Be records", *Earth Planet. Sci. Lett.*, **150**, 453–462. [DOI], [ADS] (Cited on page 57.)

Barnard, L., Lockwood, M., Hapgood, M.A., Owens, M. J., Davis, C.J. and Steinhilber, F., 2011, "Predicting space climate change", *Geophys. Res. Lett.*, **38**, L16103. [DOI], [ADS]. Online version (accessed 14 April 2013):
http://centaur.reading.ac.uk/22909/ (Cited on pages 63, 64, 65, 66, 67, and 69.)

Bartels, J., 1932, "Terrestrial-magnetic activity and its relations to solar phenomena", *Terr. Magn. Atmos. Electr.*, **37**, 1–52. [DOI], [ADS] (Cited on pages 10 and 11.)

Bartels, J., Heck, N.H. and Johnston, H.F., 1939, "The three-hour-range index measuring geomagnetic activity", *Terr. Magn. Atmos. Electr.*, **44**, 411–454. [DOI], [ADS] (Cited on pages 9 and 10.)

Bartholmes, J.C., Thielen, J., Ramos, M.H. and Gentilini, S., 2009, "The European Flood Alert System EFAS – Part 2: Statistical skill assessment of probabilistic and deterministic operational forecasts", *Hydrol. Earth Syst. Sci.*, **13**, 141–153. [DOI] (Cited on page 67.)

Beer, J., 2000, "Long-term indirect indices of solar variability", *Space Sci. Rev.*, **94**(1/2), 53–66. [DOI], [ADS] (Cited on page 57.)

Beer, J., Blinov, A., Bonani, G., Hofmann, H.J. and Finkel, R. C., 1990, "Use of ^{10}Be in polar ice to trace the 11-year cycle of solar activity", *Nature*, **347**, 164–166. [DOI], [ADS] (Cited on page 62.)

Beer, J., Tobias, S. and Weiss, N., 1998, "An active Sun throughout the Maunder minimum", *Solar Phys.*, **181**, 237–249. [DOI], [ADS] (Cited on page 62.)

Beer, J., McCracken, K. and von Steiger, R., 2012, *Cosmogenic radionuclides: theory and applications in the terrestrial and space environments*, Physics of Earth and Space Environments, 26, Springer, Berlin, Heidelberg. [DOI], [ADS] (Cited on page 57.)

Belov, A., 2000, "Large-scale modulation: view from the Earth", *Space Sci. Rev.*, **93**, 79–105. [DOI], [ADS] (Cited on page 57.)

Bhattacharyya, A and Mitra, B., 1997, "Changes in cosmic ray cut-off rigidities due to secular variations of the geomagnetic field", *Ann. Geophys.*, **15**, 734–739. [DOI], [ADS] (Cited on page 57.)

Bonino, G., Cini-Castagnoli, G., Bhandari, N. and Taricco, C., 1995, "Behavior of the heliosphere over prolonged solar quiet periods by 44Ti measurements in meteorites", *Science*, **270**, 1648–1650. [DOI], [ADS] (Cited on page 57.)

Burlaga, L.F., Behannon, K.W., Hansen, S.F., Pneuman, G.W. and Feldman, W.C., 1978, "Sources of magnetic fields in recurrent interplanetary streams", *J. Geophys. Res.*, **83**, 4177–4185. [DOI], [ADS] (Cited on page 33.)

Burton, R.K., McPherron, R.L. and Russell, C.T., 1975, "An empirical relationship between interplanetary conditions and Dst", *J. Geophys. Res.*, **80**, 4204–4214. [DOI], [ADS] (Cited on page 13.)

Caballero-Lopez, R.A.and Moraal, H., 2004, "Limitations of the force field equation to describe cosmic ray modulation", *J. Geophys. Res.*, **109**, A01101. [DOI], [ADS] (Cited on page 57.)

Cane, H.V., Wibberenz, G., Richardson, I.G. and von Rosenvinge, T.T., 1999, "Cosmic ray modulation and the solar magnetic field", *Geophys. Res. Lett.*, **29**, 565–568. [DOI], [ADS] (Cited on page 57.)

Charbonneau, P., 2010, "Dynamo Models of the Solar Cycle", *Living Rev. Solar Phys.*, **7**, lrsp-2010-3. [DOI], [ADS]. URL (accessed 14 April 2013):
http://www.livingreviews.org/lrsp-2010-3 (Cited on page 64.)

Clauer, C.R. and Kamide, Y., 1985, "DP 1 and DP 2 current systems for the March 22, 1979 substorms", *J. Geophys. Res.*, **90**, 1343–1354. [DOI], [ADS] (Cited on page 9.)

Clem, J.M. and Everson, P., 2002, "Positron Abundance in Galactic Cosmic Rays", *Astrophys. J.*, **568**, 216. [DOI], [ADS] (Cited on page 58.)

Clilverd, M.A., Clark, T.D.G., Clarke, E. and Rishbeth, H., 1998, "Increased magnetic storm activity from 1868 to 1995", *J. Atmos. Sol.-Terr. Phys.*, **60**, 1047–1056. [DOI], [ADS] (Cited on page 15.)

Clilverd, M.A., Clark, T.D.G., Clarke, E., Rishbeth, H. and Ulrich, T., 2002, "The causes of the long-term change in the aa index", *J. Geophys. Res.*, **107**(A12), A1441. [DOI], [ADS] (Cited on page 15.)

Clilverd, M.A., Clarke, E., Ulich, T., Linthe, J. and Rishbeth, H., 2005, "Reconstructing the long-term *aa* index", *J. Geophys. Res.*, **110**(A9), A07205. [DOI], [ADS] (Cited on pages 15 and 21.)

Cliver, E.W., 1994a, "Solar activity and geomagnetic storms: The first 40 years", *Eos Trans. AGU*, **75**, 569–575. [DOI], [ADS] (Cited on page 6.)

Cliver, E.W., 1994b, "Solar activity and geomagnetic storms: The corpuscular hypothesis", *Eos Trans. AGU*, **75**, 609–670. [DOI], [ADS] (Cited on page 6.)

Cliver, E.W., 1995, "Solar activity and geomagnetic storms: From M regions and flares to coronal holes and CMEs", *Eos Trans. AGU*, **76**, 75–83. [DOI], [ADS] (Cited on page 6.)

Cliver, E.W. and Ling, A.G., 2002, "Secular change in geomagnetic indices and the solar open magnetic flux during the first half of the twentieth century", *J. Geophys. Res.*, **107**(A10), A1303. [DOI], [ADS] (Cited on page 15.)

Cliver, E.W. and Ling, A.G., 2011, "The floor in the solar wind magnetic field revisited", *Solar Phys.*, **264**, 285–301. [DOI], [ADS] (Cited on pages 52, 58, and 62.)

Connick, D.E., Smith, C. and Schwadron, N.A., 2011, "Interplanetary magnetic flux depletion during protracted solar minimums", *Astrophys. J.*, **727**, 8. [DOI], [ADS] (Cited on page 61.)

Consolini, G. and De Michelis, P., 2005, "Local intermittency measure analysis of AE index: The directly driven and unloading component", *Geophys. Res. Lett.*, **32**, L05101. [DOI], [ADS] (Cited on page 9.)

Cook, G.R., Mackay, D.H. and Nandy, D., 2009, "Solar Cycle Variations of Coronal Null Points: Implications for the Magnetic Breakout Model of Coronal Mass Ejections", *Astrophys. J.*, **704**, 1021–1035. [DOI], [ADS] (Cited on page 33.)

Couzens, D.A. and King, J.H., 1986, *Interplanetary Medium Data Book – Supplement 3, 1977–1985*, NSSDC/WDC-A/R&S, 86-04, NASA, Greenbelt, MD. Online version (accessed 30 April 2013): http://www.archive.org/details/nasa_techdoc_19890001401 (Cited on page 6.)

Cowley, S.W.H., 1991, "Acceleration and heating of space plasmas: Basic concepts", *Ann. Geophys.*, **9**, 176–187. [ADS] (Cited on page 22.)

Crooker, N.U., Gosling, J.T. and Kahler, S.W., 2002, "Reducing heliospheric magnetic flux from coronal mass ejections without disconnection", *J. Geophys. Res.*, **107**(A2), SSH 3–1–SSH 3–5. [DOI], [ADS] (Cited on page 60.)

Crooker, N.U., Kahler, S.W., Larson, D.E. and Lin, R.P., 2004, "Large-scale magnetic field inversions at sector boundaries", *J. Geophys. Res.*, **109**, A03108. [DOI], [ADS] (Cited on pages 40 and 41.)

Davis, T.N. and Sugiura, M., 1966, "Auroral electrojet activity index AE and its universal time variations", *J. Geophys. Res.*, **71**, 785–801. [DOI], [ADS] (Cited on page 7.)

Dungey, J.W., 1961, "Interplanetary Magnetic Field and the Auroral Zones", *Phys. Rev. Lett.*, **6**, 47–48. [DOI], [ADS] (Cited on page 13.)

Erdős, G. and Balogh, A., 2012, "Magnetic Flux Density Measured in Fast and Slow Solar Wind Streams", *Astrophys. J.*, **753**, 130. [DOI], [ADS] (Cited on page 41.)

Evenson, P., 1998, "Cosmic Ray Electrons", *Space Sci. Rev.*, **83**, 63–73. [DOI], [ADS] (Cited on page 57.)

Feynman, J. and Crooker, N.U., 1978, "The solar wind at the turn of the century", *Nature*, **275**, 626–627. [DOI], [ADS] (Cited on page 20.)

Finch, I.D., 2008, *The Use of Geomagnetic Activity Observations in Studies of Solar Wind – Magnetosphere Coupling and Centennial Solar Change*, Ph.D. thesis, Southampton University, Southampton, UK. Online version (accessed 14 April 2013): http://www.eiscat.rl.ac.uk/Members/mike/IvansThesis/Ivan_thesis_final.pdf/file_view (Cited on pages 12 and 18.)

Finch, I.D. and Lockwood, M., 2007, "Solar wind-magnetosphere coupling functions on timescales of 1 day to 1 year", *Ann. Geophys.*, **25**, 495–506. [DOI], [ADS] (Cited on pages 6, 13, 22, 23, and 24.)

Finch, I.D., Lockwood, M. and Rouillard, A.P., 2008, "Effects of solar wind magnetosphere coupling recorded at different geomagnetic latitudes: Separation of directly-driven and storage/release systems", *Geophys. Res. Lett.*, **35**, L21105. [DOI], [ADS]. Online version (accessed 14 April 2013): http://centaur.reading.ac.uk/7510/ (Cited on pages 11, 12, 27, and 28.)

Fisk, L.A., 1999, "An overview of the transport of galactic and anomalous cosmic rays in the heliopshere: theory", *Adv. Space Res.*, **23**, 415–423. [DOI], [ADS] (Cited on page 60.)

Fisk, L.A. and Schwadron, N.A, 2001, "The behaviour of the open magnetic flux of the Sun", *Astrophys. J.*, **560**, 425–438. [DOI], [ADS] (Cited on page 60.)

Fitzenreiter, R.J., Ogilvie, K.W., Chornay, D.J. and Keller, J., 1998, "Observations of electron velocity distribution functions in the solar wind by the WIND Spacecraft: High angular resolution Strahl measurements", *Geophys. Res. Lett.*, **25**, 249–252. [DOI], [ADS] (Cited on page 60.)

Forbush, S.E., 1958, "Cosmic ray intensity variations during two solar cycles", *J. Geophys. Res.*, **63**, 651–669. [DOI], [ADS] (Cited on page 57.)

Foullon, C., Lavraud, B., Luhmann, J.G. et al., 2011, "Plasmoid Releases In The Heliospheric Current Sheet And Associated Coronal Hole Boundary Layer Evolution", *Astrophys. J.*, **737**, 16. [DOI], [ADS] (Cited on pages 40 and 41.)

Gleeson, L.J. and Axford, W.I., 1968, "Solar Modulation of Galactic Cosmic Rays", *Astrophys. J.*, **154**, 1011–1026. [DOI], [ADS] (Cited on page 57.)

Gleissberg, W., 1944, "A table of secular variations of the solar cycle", *J. Geophys. Res.*, **49**, 243–244. [DOI], [ADS] (Cited on page 20.)

Graham, G., 1724a, "An Account of Observations Made of the Variation of the Horizontal Needle at London, in the Latter Part of the Year 1722, and Beginning of 1723", *Philos. Trans. R. Soc. London*, **33**, 96–107. [DOI], [ADS] (Cited on page 5.)

Graham, G., 1724b, "Observations of the Dipping Needle, Made at London, in the Beginning of the Year 1723", *Philos. Trans. R. Soc. London*, **33**, 332–339. [DOI], [ADS] (Cited on page 5.)

Hapgood, M.A., 1992, "Space physics coordinate transformations: A user guide", *Planet. Space Sci.*, **40**, 711–717. [DOI], [ADS] (Cited on page 13.)

Hapgood, M.A., 1993, "A double solar-cycle variation in the 27-day recurrence of geomagnetic activity", *Ann. Geophys.*, **11**, 248–253. [ADS] (Cited on page 48.)

Hapgood, M.A., Lockwood, M., Bowe, G.A., Willis, D.M. and Tulunay, Y.K., 1991, "Variability of the interplanetary medium at 1 a.u. over 24 years: 1963–1986", *Planet. Space Sci.*, **39**, 411–419. [DOI], [ADS] (Cited on page 13.)

Harrison, R.G., 2004, "Long-term measurements of the global atmospheric electric circuit at Eskdalemuir, Scotland, 1911–1981", *Atmos. Res.*, **70**, 1–19. [DOI], [ADS] (Cited on page 21.)

Hathaway, D.H., 2010, "The Solar Cycle", *Living Rev. Solar Phys.*, **7**, lrsp-2010-1. [DOI], [ADS]. URL (accessed 14 April 2013):
http://www.livingreviews.org/lrsp-2010-1 (Cited on pages 51 and 66.)

Heber, B., Ferrando, P., Raviart, A. et al., 1999, "Differences in the temporal variations of galactic cosmic ray electrons and protons: Implications from Ulysses at solar minimum", *Geophys. Res. Lett.*, **26**(14), 2133–2136. [DOI], [ADS] (Cited on page 58.)

Hoeksema, J.T., Wilcox, J.M. and Scherrer, P.H., 1982, "Structure of the heliospheric current sheet in the early portion of Sunspot Cycle 21", *J. Geophys. Res.*, **87**, 10 331–10 338. [DOI], [ADS] (Cited on page 33.)

Hong, X. and Billings, S.A., 1999, "Time series multistep-ahead predictability estimation and ranking", *J. Forecast.*, **18**, 139–149. [DOI] (Cited on page 63.)

Hoyt, D. and Schatten, K.A., 1998, "Group sunspot numbers: a new solar activity reconstruction", *Solar Phys.*, **181**, 491–512. [DOI], [ADS] (Cited on page 51.)

Jankowski, J. and Sucksdorff, C., 1996, *Guide for Magnetic Measurements and Observatory Practice*, IAGA, Warsaw; Boulder, CO. Online version (accessed 14 April 2013): http://iugg.org/IAGA/iaga_pages/pubs_prods/Guides.htm (Cited on page 6.)

Jardine, M.M., Wood, K., Collier-Cameron, A., Donati, J.-F. and Mackay, D.H., 2002, "Inferring X-ray coronal structures from Zeeman-Doppler images", *Mon. Not. R. Astron. Soc.*, **336**, 1364–1370. [DOI], [ADS] (Cited on page 33.)

Jarvis, M.J., 2004, "Observed tidal variation in the lower thermosphere through the 20th century and the possible implication of ozone depletion", *J. Geophys. Res.*, **110**, A04303. [DOI], [ADS] (Cited on page 17.)

Jokipii, J.R., 1991, "Variations of the cosmic ray flux with time", in *The Sun in Time*, (Eds.) Sonnet, C.P., Giampapa, M.S., Matthews, M.S., pp. 205–221, University of Arizona Press, Tucson, AZ (Cited on page 57.)

Jokipii, J.R., Levy, E.H. and Hubbard, W.B., 1977, "Effects of particle drift on cosmic ray transport. I. General properties, application to solar modulation", *Astrophys. J.*, **213**, 861–868. [DOI], [ADS] (Cited on page 57.)

Jones, G.H., Balogh, A. and Forsyth, R.J., 1998, "Radial heliospheric magnetic fields detected by Ulysses", *Geophys. Res. Lett.*, **25**, 3109–3112. [DOI], [ADS] (Cited on page 36.)

Jones, G.H., Balogh, A. and Smith, E.J., 2003, "Solar magnetic field reversal as seen at Ulysses", *Geophys. Res. Lett.*, **30**, 8028. [DOI], [ADS] (Cited on page 60.)

Kahler, S.W. and Lin, R.P., 1995, "An examination of directional discontinuities and magnetic polarity changes around interplanetary sector boundaries using $E > 2$ keV electrons", *Solar Phys.*, **161**, 183–195. [DOI], [ADS] (Cited on page 40.)

Kahler, S.W., Crooker, N.U. and Gosling, J.T., 1996, "The topology of intrasector reversals of the interplanetary magnetic field", *J. Geophys. Res.*, **101**, 24 373–24 382. [DOI], [ADS] (Cited on page 38.)

Kahler, S.W., Crooker, N.U. and Gosling, J.T., 1998, "Properties of interplanetary magnetic sector boundaries based on electron heat-flux flow directions", *J. Geophys. Res.*, **103**, 20,603–20,612. [DOI], [ADS] (Cited on page 40.)

Karlsson, S.B.P., Opgenoorth, H.J., Eglitis, P. et al., 2000, "Solar wind control of magnetospheric energy content: substorm quenching and multiple onsets", *J. Geophys. Res.*, **105**, 5335–5356. [DOI], [ADS] (Cited on page 30.)

King, J.H. and Papitashvili, N.E., 2005, "Solar wind spatial scales in and comparisons of hourly Wind and ACE plasma and magnetic field data", *J. Geophys. Res.*, **110**, A02104. [DOI], [ADS] (Cited on page 6.)

Kozyra, J.U. and Liemohn, M.W., 2003, "Ring Current Energy Input and Decay", *Space Sci. Rev.*, **109**, 105–131. [DOI], [ADS] (Cited on page 13.)

Krzysztofowicz, R., 1998, "Probabilistic Hydrometeorological Forecasts: Toward a New Era in Operational Forecasting", *Bull. Am. Meteorol. Soc.*, **79**, 243–251. [DOI], [ADS] (Cited on page 67.)

Larson, D.E., Lin, R.P., Ergun, R.E. et al., 1997, "Using energetic electrons to probe the topology of the October 18-20, 1995 magnetic cloud", *Adv. Space Res.*, **20**, 655. [DOI], [ADS] (Cited on page 60.)

Lean, J., Wang, Y.-M. and Sheeley Jr, N.R., 2002, "The effect of increasing solar activity on the Sun's total and open magnetic flux during multiple cycles: Implications for solar forcing of climate", *Geophys. Res. Lett.*, **29**(24), 2224. [DOI], [ADS] (Cited on page 60.)

Lockwood, M., 1997, "Solar Wind - Magnetosphere Coupling", in *Incoherent Scatter: Theory, Practice and Science*, Collection of lectures given in Cargese, Corsica, 1995, (Ed.) Alcaydé, D., Technical Report, 97/53, pp. 121–186, EISCAT Scientific Association, Kiruna, Sweden. Online version (accessed 14 April 2013): http://www.eiscat.com/groups/Documentation/BlueBooks/Cargesebook.pdf/view (Cited on page 8.)

Lockwood, M., 2001, "Long-Term Variations in the Magnetic Fields of the Sun and the Heliosphere: their origin, effects and implications", *J. Geophys. Res.*, **106**, 16 021–16 038. [DOI], [ADS] (Cited on pages 15 and 57.)

Lockwood, M., 2003, "Twenty-three cycles of changing open solar flux", *J. Geophys. Res.*, **108**, 1128. [DOI], [ADS] (Cited on pages 15, 51, 57, 61, and 64.)

Lockwood, M., 2004, "Solar Outputs, their variations and their effects of Earth", in *The Sun, Solar Analogs and the Climate*, (Eds.) Rüedi, I., Güdel, M., Schmutz, W., Saas-Fee Advanced Courses, 34, pp. 107–304, Springer, Berlin; New York (Cited on pages 22, 33, and 60.)

Lockwood, M., 2010, "Solar change and climate: an update in the light of the current exceptional solar minimum", *Proc. R. Soc. London, Ser. A*, **466**, 303–329. [DOI], [ADS] (Cited on pages 15, 53, 63, 64, 65, and 67.)

Lockwood, M. and Fröhlich, C., 2007, "Recent oppositely directed trends in solar climate forcings and the global mean surface air temperature", *Proc. R. Soc. London, Ser. A*, **463**, A09109. [DOI], [ADS] (Cited on page 64.)

Lockwood, M. and Owens, M.J., 2009, "The Accuracy of Using the Ulysses Result of the Spatial Invariance of the Radial Heliospheric Field to Compute the Open Solar Flux", *Astrophys. J.*, **701**, 964–973. [DOI], [ADS] (Cited on page 36.)

Lockwood, M. and Owens, M.J., 2011, "Centennial changes in the heliospheric magnetic field and open solar flux: The consensus view from geomagnetic data and cosmogenic isotopes and its implications", *J. Geophys. Res.*, **116**, A04109. [DOI] (Cited on pages 46, 47, 49, 50, and 58.)

Lockwood, M. and Owens, M.J., 2013, "Comment on 'What causes the flux excess in the heliospheric magnetic field?' by E.J. Smith", *J. Geophys. Res.*, **118**. [DOI] (Cited on pages 34, 38, and 41.)

Lockwood, M., Stamper, R. and Wild, M.N., 1999a, "A doubling of the Sun's coronal magnetic field during the last 100 years", *Nature*, **399**, 437–439. [DOI], [ADS] (Cited on pages 20, 42, 46, 47, 48, 50, and 57.)

Lockwood, M, Stamper, R., Wild, M.N., Balogh, A. and Jones, G., 1999b, "Our changing Sun", *Astron. Geophys.*, **40**, 4.10–4.16. [DOI], [ADS] (Cited on page 33.)

Lockwood, M., Forsyth, R.B., Balogh, A. and McComas, D.J., 2004, "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", *Ann. Geophys.*, **22**, 1395–1405. [DOI], [ADS] (Cited on pages 33 and 34.)

Lockwood, M., Rouillard, A.P., Finch, I. and Stamper, R., 2006a, "Comment on 'The IDV index: Its derivation and use in inferring long-term variations of the interplanetary magnetic field strength' by Leif Svalgaard and Edward W. Cliver", *J. Geophys. Res.*, **111**, A09109. [DOI], [ADS]. Online version (accessed 14 April 2013): http://www.eiscat.rl.ac.uk/Members/mike/publications/ (Cited on pages 34 and 42.)

Lockwood, M., Whiter, D., Hancock, B., Henwood, R., Ulich, T., Linthe, H.J., Clarke, E. and Clilverd, M.A., 2006b, "The long-term drift in geomagnetic activity: calibration of the aa index using data from a variety of magnetometer stations", *Ann. Geophys.*, submitted. Online version (accessed 14 April 2013): http://www.eiscat.rl.ac.uk/Members/mike/publications/ (Cited on pages 11, 12, 15, 16, 17, 18, 21, and 37.)

Lockwood, M., Hairston, M.R., Finch, I.D. and Rouillard, A.P., 2009a, "Transpolar voltage and polar cap flux during the substorm cycle and steady convection events", *J. Geophys. Res.*, **114**, A01210. [DOI], [ADS]. Online version (accessed 14 April 2013):
http://centaur.reading.ac.uk/7224/ (Cited on page 27.)

Lockwood, M., Owens, M. and Rouillard, A.P., 2009b, "Excess open solar magnetic flux from satellite data: 1. Analysis of the third perihelion Ulysses pass", *J. Geophys. Res.*, **114**, A11103. [DOI], [ADS] (Cited on pages 33, 34, 35, 36, and 38.)

Lockwood, M., Owens, M. and Rouillard, A.P., 2009c, "Excess open solar magnetic flux from satellite data: 2. A survey of kinematic effects", *J. Geophys. Res.*, **114**, A11104. [DOI], [ADS] (Cited on pages 34, 36, 37, 39, 40, 49, and 60.)

Lockwood, M., Rouillard, A.P. and Finch, I.D., 2009d, "The rise and fall of open solar flux during the current grand solar maximum", *Astrophys. J.*, **700**(2), 937–944. [DOI], [ADS] (Cited on pages 12, 30, 31, 32, 42, 46, 47, 49, 64, and 66.)

Lockwood, M., Harrison, R.G., Owens, M.J., Barnard, L., Woollings, T. and Steinhilber, F., 2011a, "The solar influence on the probability of relatively cold UK winters in the future", *Environ. Res. Lett.*, **6**, 034004. [DOI], [ADS] (Cited on pages 65 and 66.)

Lockwood, M., Owens, M. J., Barnard, L., Davis, C. J. and Steinhilber, F., 2011b, "The persistence of solar activity indicators and the descent of the Sun into Maunder Minimum conditions", *Geophys. Res. Lett.*, **38**, L22105. [DOI], [ADS] (Cited on page 63.)

Lockwood, M., Owens, M.J., Barnard, L., Davis, C.J. and Steinhilber, F., 2011c, "The persistence of solar activity indicators and the descent of the Sun into Maunder Minimum conditions", *Geophys. Res. Lett.*, **38**, L22105. [DOI], [ADS]. Online version (accessed 14 April 2013):
http://centaur.reading.ac.uk/25495/ (Cited on page 52.)

Lockwood, M., Owens, M.J., Barnard, L., Davis, C.J. and Thomas, S., 2012, "Solar cycle 24: What is the Sun up to?", *Astron. Geophys.*, **53**, 3.09–3.15. [DOI], [ADS] (Cited on pages 53 and 66.)

Lockwood, M., Barnard, L., Nevanlinna, H., Owens, M.J., Harrison, R.G., Rouillard, A.P. and Davis, C.J., 2013a, "Reconstruction of geomagnetic activity and near-Earth interplanetary conditions over the past 167 years: 1. A new geomagnetic data composite", *Ann. Geophys.*, submitted (Cited on pages 10, 11, 14, 20, 21, 53, and 56.)

Lockwood, M., Barnard, L., Nevanlinna, H., Owens, M.J., Harrison, R.G., Rouillard, A.P. and Davis, C.J., 2013b, "Reconstruction of geomagnetic activity and near-Earth interplanetary conditions over the past 167 years: 2. A new reconstruction of the interplanetary magnetic field", *Ann. Geophys.*, submitted (Cited on pages 47, 53, and 56.)

Low, B.C., 2001, "Coronal mass ejections, magnetic flux ropes, and solar magnetism", *J. Geophys. Res.*, **106**, 25,141–25,164. [DOI], [ADS] (Cited on page 60.)

Lu, H., Li, Y., Clilverd, M.A. and Jarvis, M.J., 2012, "Trend and abrupt changes in long-term geomagnetic indices", *J. Geophys. Res.*, **117**, A05318. [DOI], [ADS] (Cited on page 15.)

Luhmann, J.G., Solomon, S.C., Linker, J.A., Lyon, J., Mikic, Z., Odstrčil, D., Wang, W. and Wiltberger, M., 2004, "Coupled model simulation of a Sun-to-Earth space weather event", *J. Atmos. Sol.-Terr. Phys.*, **66**, 1243–1256. [DOI], [ADS] (Cited on page 33.)

Mackay, D.H and Lockwood, M., 2002, "The Evolution of the Sun's Open magnetic Flux: II. Full solar cycle simulations", *Solar Phys.*, **209**(2), 287–309. [ADS] (Cited on page 60.)

Mackay, D.H. and Yeates, A.R., 2012, "The Sun's Global Photospheric and Coronal Magnetic Fields: Observations and Models", *Living Rev. Solar Phys.*, **9**, lrsp-2012-6. [DOI], [ADS], [arXiv:1211.6545]. URL (accessed 14 April 2013):
http://www.livingreviews.org/lrsp-2012-6 (Cited on page 33.)

MacMillan, S. and Clarke, E., 2011, "Resolving issues concerning Eskdalemuir geomagnetic hourly values", *Ann. Geophys.*, **29**, 283–288. [DOI], [ADS] (Cited on page 21.)

Marsch, E., 1999, "Solar wind models from the Sun to 1 AU: Constraints by in situ and remote sensing measurements", *Space Sci. Rev.*, **87**, 1–24. [DOI], [ADS] (Cited on page 33.)

Martini, D., Linthe, H.-J., Pandey, V.S. and Lee, D.-H., 2012, "Correcting the geomagnetic IHV index of the Eskdalemuir observatory", *J. Geophys. Res.*, **117**, A06211. [DOI], [ADS] (Cited on page 17.)

Masarik, J. and Beer, J., 1999, "Simulation of particle fluxes and cosmogenic nuclide production in the Earth's atmosphere", *J. Geophys. Res.*, **104**(13), 12,099–12,112. [DOI], [ADS] (Cited on page 57.)

Mayaud, P.-N., 1971, "Une mesure planétaire d'activité magnetique, basée sur deux observatoires antipodaux", *Ann. Geophys. France*, **27**, 67–70 (Cited on pages 9, 10, and 15.)

Mayaud, P.-N., 1972, "The *aa* indices: A 100T-year series characterizing the magnetic activity", *J. Geophys. Res.*, **77**, 6870–6874. [DOI], [ADS] (Cited on pages 9 and 15.)

Mayaud, P.-N., 1980, *Derivation, Meaning and Use of Geomagnetic Indices*, Geophysical Monograph, 22, American Geophysical Union, Washington, DC. [DOI], [ADS] (Cited on pages 9 and 15.)

McCracken, K.G., 2007, "Heliomagnetic field near Earth, 1428–2005", *J. Geophys. Res.*, **112**, A09106. [DOI], [ADS] (Cited on page 57.)

McCracken, K.G. and Beer, J., 2007, "Long-term changes in the cosmic ray intensity at Earth, 1428–2005", *J. Geophys. Res.*, **112**, A10101. [DOI], [ADS] (Cited on page 57.)

McCracken, K.G. and McDonald, F.B., 2001, "The long-term modulation of the galactic cosmic radiation, 1500-2000", in *Proceedings of ICRC 2001*, 27th International Cosmic Ray Conference, (Ed.) Schlickeiser, R., 27, pp. 3753–3756, Copernicus, Katlenburg-Lindau, Germany. [ADS] (Cited on page 57.)

McDonald, F.B., Lal, N. and McGuire, R.E., 1993, "The role of drifts and global merged interaction regions in the long-term modulation of cosmic rays", *J. Geophys. Res.*, **98**, 1243–1256. [DOI], [ADS] (Cited on page 57.)

McPherron, R.L., Russell, C.T. and Aubry, M., 1973, "Satellite studies of magnetospheric substorms on August 15, 1978, 9, Phenomenological model for substorms", *J. Geophys. Res.*, **78**, 3131–3149. [DOI], [ADS] (Cited on page 27.)

Mursula, K. and Martini, D., 2006, "Centennial increase in geomagnetic activity: Latitudinal differences and global estimates", *J. Geophys. Res.*, **111**, A08209. [DOI], [ADS] (Cited on page 16.)

Mursula, K. and Martini, D., 2007, "A new verifiable measure of centennial geomagnetic activity: Modifying the K index method for hourly data", *Geophys. Res. Lett.*, **34**, L22107. [DOI], [ADS] (Cited on page 15.)

Mursula, K., Martini, D. and Karinen, A., 2004, "Did Open Solar Magnetic Field Increase During The Last 100 Years? A Reanalysis of Geomagnetic Activity", *Solar Phys.*, **224**, 85–94. [DOI], [ADS] (Cited on pages 16 and 21.)

Muscheler, R., Joos, F., Müller, S.A. and Snowball, I., 2005, "How unusual is today's solar activity?", *Nature*, **436**, E3–E4. [DOI], [ADS] (Cited on page 59.)

Muscheler, R., Joos, F., Beer, J., Müller, S.A., Vonmoos, M. and Snowball, I., 2007, "Solar activity during the last 1000 yr inferred from radionuclide records", *Quat. Sci. Rev.*, **26**, 82–97. [DOI], [ADS] (Cited on page 57.)

Neher, H.V., Peterson, V.Z. and Stern, E.A., 1953, "Fluctuations and latitude effect of cosmic rays at high altitudes and latitudes", *Phys. Rev.*, **90**, 655–674. [DOI], [ADS] (Cited on page 57.)

Nelder, J.A. and Mead, R., 1965, "A simplex method for function minimization", *Comput. J.*, **7**, 308–313. [DOI] (Cited on pages 30 and 54.)

Neugebauer, M., Forsyth, R.J., Galvin, A.B. et al., 1998, "Spatial structure of the solar wind and comparisons with solar data and models", *J. Geophys. Res.*, **103**, 14,587–14,600. [DOI], [ADS] (Cited on page 33.)

Nevanlinna, H., 2004, "Results of the Helsinki magnetic observatory 1844–1912", *Ann. Geophys.*, **22**, 1691–1704. [DOI], [ADS] (Cited on pages 51 and 56.)

Nevanlinna, H. and Häkkinen, L., 2010, "Results of Russian geomagnetic observatories in the 19th century: magnetic activity, 1841–1862", *Ann. Geophys.*, **28**, 917–926. [DOI], [ADS] (Cited on page 56.)

Nevanlinna, H. and Kataja, E., 1993, "An extension to the geomagnetic activity index series *aa* for two solar cycles (1844–1868)", *Geophys. Res. Lett.*, **20**, 2703–2706. [DOI], [ADS] (Cited on page 51.)

Nishida, A., 1968, "Coherence of geomagnetic *DP*2 fluctuations with interplanetary magnetic variations", *J. Geophys. Res.*, **73**, 5549–5559. [DOI], [ADS] (Cited on page 13.)

O'Brien, K., 1979, "Secular variations in the production of cosmogenic isotopes in the Earth's atmosphere", *J. Geophys. Res.*, **84**, 423–431. [DOI], [ADS] (Cited on page 57.)

O'Brien, K., de la Zerda, A., Shea, M.A. and Smart, D.F., 1991, "The production of cosmogenic isotopes in the Earth's atmosphere and their inventories", in *The Sun in Time*, (Eds.) Sonnet, C.P., Giampapa, M.S., Matthews, M.S., pp. 317–342, University of Arizona Press, Tucson, AZ. [ADS] (Cited on page 57.)

Owens, M.J. and Crooker, N.U., 2006, "Coronal mass ejections and magnetic flux buildup in the heliosphere", *J. Geophys. Res.*, **111**, A10104. [DOI], [ADS] (Cited on page 60.)

Owens, M.J. and Crooker, N.U., 2007, "Reconciling the electron counterstreaming and dropout occurrence rates with the heliospheric flux budget", *J. Geophys. Res.*, **112**, A06106. [DOI], [ADS] (Cited on page 60.)

Owens, M.J. and Forsyth, R.J., 2013, "The Heliospheric Magnetic Field", *Living Rev. Solar Phys.*, submitted (Cited on page 60.)

Owens, M.J. and Lockwood, M., 2012, "Cyclic loss of open solar flux since 1868: The link to heliospheric current sheet tilt and implications for the Maunder Minimum", *J. Geophys. Res.*, **117**, A04102. [DOI], [ADS] (Cited on pages 61 and 62.)

Owens, M.J., Schwadron, N.A., Crooker, N.U., Hughes, W.J. and Spence, H.E., 2007, "Role of coronal mass ejections in the heliospheric Hale cycle", *Geophys. Res. Lett.*, **34**, L06104. [DOI], [ADS] (Cited on page 60.)

Owens, M.J., Arge, C.N., Crooker, N.U., Schwadron, N.A. and Horbury, T.S., 2008a, "Estimating total heliospheric magnetic flux from single-point in situ measurements", *J. Geophys. Res.*, **113**, A12103. [DOI], [ADS] (Cited on pages 34 and 37.)

Owens, M.J., Crooker, N.U. and Schwadron, N.A., 2008b, "Suprathermal electron evolution in a Parker spiral magnetic field", *J. Geophys. Res.*, **113**, A11104. [DOI], [ADS] (Cited on page 60.)

Owens, M.J., Crooker, N.U. and Lockwood, M., 2011a, "How is open solar magnetic flux lost over the solar cycle?", *J. Geophys. Res.*, **116**, A04111. [DOI], [ADS] (Cited on page 61.)

Owens, M.J., Lockwood, M., Barnard, L. and Davis, C.J., 2011b, "Solar cycle 24: Implications for energetic particles and long-term space climate change", *Geophys. Res. Lett.*, **38**, L19106. [DOI], [ADS] (Cited on pages 66 and 68.)

Owens, M.J., Usoskin, I.G. and Lockwood, M., 2012, "Heliospheric modulation of galactic cosmic rays during grand solar minima: Past and future variations", *Geophys. Res. Lett.*, **39**, L19102. [DOI], [ADS]. Online version (accessed 14 April 2013):
http://cc.oulu.fi/~usoskin/personal/2012GL053151.pdf (Cited on page 62.)

Owens, M.J., Crooker, N.U. and Lockwood, M., 2013, "Solar origin of heliospheric magnetic field inversions: Evidence for coronal magnetic loop opening within pseudostreamers", *J. Geophys. Res.*, **118**, 1868–1879. [DOI] (Cited on pages 33, 40, and 41.)

Parker, E.N., 1958, "Dynamics of the interplanetary gas and magnetic fields", *Astrophys. J.*, **128**, 664–676. [DOI], [ADS] (Cited on page 26.)

Parker, E.N., 1963, *Interplanetary Dynamical Processes*, Interscience Monographs and Texts in Physics and Astronomy, 8, Interscience Publishers, New York. [ADS] (Cited on page 26.)

Parker, E.N., 1965, "The passage of energetic charged particles through interplanetary space", *Planet. Space Sci.*, **13**, 9. [DOI], [ADS] (Cited on page 57.)

Petrovay, K., 2010, "Solar Cycle Prediction", *Living Rev. Solar Phys.*, **7**, lrsp-2010-6. [DOI], [ADS], [arXiv:1012.5513]. URL (accessed 14 April 2013):
http://www.livingreviews.org/lrsp-2010-6 (Cited on page 63.)

Potgieter, M.S., 1998, "The modulation of galactic cosmic rays in the heliosphere: theory and models", *Space Sci. Rev.*, **83**, 147–158. [DOI], [ADS] (Cited on page 57.)

Potgieter, M.S., 2013, "Solar Modulation of Cosmic Rays", *Living Rev. Solar Phys.*, **10**, lrsp-2013-3. [DOI], [ADS], [arXiv:1306.4421]. URL (accessed 14 July 2013):
http://www.livingreviews.org/lrsp-2013-3 (Cited on page 57.)

Pulkkinen, T., 2007, "Space Weather: Terrestrial Perspective", *Living Rev. Solar Phys.*, **4**, lrsp-2007-1. [DOI], [ADS]. URL (accessed 14 April 2013):
http://www.livingreviews.org/lrsp-2007-1 (Cited on pages 13 and 27.)

Pulkkinen, T.I., Nevanlinna, H., Pulkkinen, P.J. and Lockwood, M., 2001, "The Earth-Sun connection in time scales from years to decades to centuries", *Space Sci. Rev.*, **95**(1/2), 625–637. [DOI], [ADS] (Cited on page 20.)

Raisbeck, G.M. and Yiou, F., 2004, "Comment on 'Millennium Scale Sunspot Number Reconstruction: Evidence for an Unusually Active Sun Since the 1940s"', *Phys. Rev. Lett.*, **92**, 199001. [DOI], [ADS] (Cited on page 59.)

Richter, P.H., 1995, *Estimating Errors in Least-Squares Fitting*, TDA Progress Report, 42-122, JPL, Pasadena, CA. [ADS]. Online version (accessed 14 April 2013):
http://tmo.jpl.nasa.gov/progress_report/42-122/title.htm (Cited on page 46.)

Riley, P. and Gosling, J.T., 2007, "On the origin of near-radial magnetic fields in the heliosphere: Numerical simulations", *J. Geophys. Res.*, **112**, A06115. [DOI], [ADS] (Cited on page 36.)

Riley, P., Ben-Nun, M., Linker, J.A. et al., 2013, "A Multi-Observatory Inter-Comparison of Line-of-Sight Synoptic Solar Magnetograms", *Solar Phys.*, **286**. [DOI] (Cited on page 36.)

Rouillard, A. and Lockwood, M., 2004, "Oscillations in the open solar magnetic flux with period 1.68 years: imprint on galactic cosmic rays and implications for heliospheric shielding", *Ann. Geophys.*, **22**, 4381–4396. [DOI] (Cited on page 57.)

Rouillard, A.P., Lockwood, M. and Finch, I.D., 2007, "Centennial changes in the solar wind speed and in the open solar flux", *J. Geophys. Res.*, **112**, A05103. [DOI], [ADS] (Cited on pages 12, 30, 31, 33, 42, 46, 47, 48, and 49.)

Rouillard, A.P, Davies, J.A., Lavraud, B. et al., 2010, "Intermittent release of transients in the slow solar wind: 1. Remote sensing observations", *J. Geophys. Res.*, **115**, A04103. [DOI], [ADS] (Cited on page 33.)

Russell, C.T., 1971, "Geophysical coordinate transformations", *Cosmic Electrodyn.*, **2**, 184–196. [ADS]. Online version (accessed 14 April 2013):
http://www-ssc.igpp.ucla.edu/personnel/russell/papers/gct1.html/ (Cited on page 13.)

Russell, C.T., Luhmann, J.G. and Jian, L.K., 2010, "How unprecedented a solar minimum?", *Rev. Geophys.*, **48**, RG2004. [DOI], [ADS]. Online version (accessed 14 April 2013):
http://aten.igpp.ucla.edu/personnel/russell/papers/how_unprecedented_solar_minimum.pdf
(Cited on pages 15 and 53.)

Sabine, E, 1851, "On Periodical Laws Discoverable in the Mean Effects of the Larger Magnetic Disturbances", *Philos. Trans. R. Soc. London*, **141**, 123–139. [ADS] (Cited on page 5.)

Sabine, E, 1852, "On Periodical Laws discoverable in the mean effects of the larger Magnetic Disturbances. – No. II", *Philos. Trans. R. Soc. London*, **142**, 103–124. [ADS] (Cited on page 5.)

Sargent, H.H., 1986, "The 27-day recurrence index", in *Solar Wind-Magnetosphere Coupling*, Proceedings of the Chapman Conference, Pasadena, CA, February 12 – 15, 1985, (Eds.) Kamide, Y., Slavin, J.A., pp. 143–148, Terra Scientific; Reidel, Tokyo; Dordrecht. [ADS] (Cited on pages 20 and 46.)

Schatten, K.H., 1999, "Models for Coronal and Interplanetary magnetic fields: a critical commentary", in *Sun-Earth plasma connections*, (Eds.) Burch, J.L., Carovillano, R.L., Antiochos, S.K., Geophysical Monograph, 109, pp. 129–142, American Geophysical Union, Washington, DC (Cited on page 32.)

Schatten, K.H., Wilcox, J.M. and Ness, N.F., 1969, "A model of interplanetary and coronal magnetic fields, PFSS method", *Solar Phys.*, **6**, 442–455. [DOI], [ADS] (Cited on page 32.)

Schatten, K.H., Scherrer, P.H., Svalgaard, L. and Wilcox, J.M., 1978, "Using dynamo theory to predict the sunspot number during solar cycle 21", *Geophys. Res. Lett.*, **5**, 411–414. [DOI], [ADS] (Cited on page 72.)

Schrijver, C.J., DeRosa, M.L. and Title, A.M., 2002, "What is missing from our understanding of long-term solar and Heliospheric activity?", *Astrophys. J.*, **577**, 1006–1012. [DOI], [ADS] (Cited on page 60.)

Schwadron, N.A., Owens, M.J. and Crooker, N.U., 2008, "The Heliospheric Magnetic Field over the Hale Cycle", *Astrophys. Space Sci. Trans.*, **4**, 19–26. [DOI], [ADS] (Cited on page 60.)

Sheeley Jr, N.R., 2005, "Surface Evolution of the Sun's Magnetic Field: A Historical Review of the Flux-Transport Mechanism", *Living Rev. Solar Phys.*, **2**, lrsp-2005-5. [DOI], [ADS]. URL (accessed 14 April 2013):
http://www.livingreviews.org/lrsp-2005-5 (Cited on page 60.)

Sheeley Jr, N.R. and Wang, Y.-M., 2001, "Coronal inflows and sector magnetism", *Astrophys. J.*, **562**, L107–L110. [DOI], [ADS] (Cited on page 61.)

Sheeley Jr, N.R., Harvey, J.W. and Feldman, W.C., 1976, "Coronal holes, solar wind streams and recurrent geomagnetic disturbances 1973 - 1976", *Solar Phys.*, **49**, 271–278. [DOI], [ADS] (Cited on page 46.)

Shi, X-F., Zong, Q-G. and Wang, Y-F., 2012, "Comparison between the ring current energy injection and decay under southward and northward IMF Bz conditions during geomagnetic storms", *Sci. China Tech. Sci.*, **55**, 2769–2777. [DOI] (Cited on page 13.)

Simpson, J.A., 2000, "The cosmic ray nucleonic component: the invention and scientific uses of the neutron monitor", *Space Sci. Rev.*, **93**, 11–32. [DOI], [ADS] (Cited on page 57.)

Smith, E.J., 2011, "What causes the flux excess in the heliospheric magnetic field?", *J. Geophys. Res.*, **116**, A12101. [DOI] (Cited on pages 33 and 38.)

Smith, E.J., 2013, "Reply to comment by M. Lockwood and M. J. Owens on 'What causes the flux excess in the heliospheric magnetic field?"', *J. Geophys. Res.*, **118**, 1888–1890. [DOI] (Cited on pages 33 and 38.)

Smith, E.J. and Balogh, A., 1995, "Ulysses observations of the radial magnetic field", *Geophys. Res. Lett.*, **22**, 3317–3320. [ADS] (Cited on page 33.)

Smith, E.J. and Balogh, A., 2003, "Open magnetic flux: Variation with latitude and solar cycle", in *Solar Wind Ten*, Proceedings of the Tenth International Solar Wind Conference, Pisa, Italy, 17 – 21 June 2002, (Eds.) Velli, M., Bruno, R, Malara, F., AIP Conference Series, 679, pp. 67–70, American Institute of Physics, Melville, NY. [DOI], [ADS] (Cited on page 33.)

Smith, E.J., Balogh, A., Forsyth, R.J. and McComas, D.J., 2001, "Ulysses in the south polar cap at solar maximum", *Geophys. Res. Lett.*, **28**, 4159–4162. [DOI], [ADS] (Cited on page 33.)

Smith, E.J., Marsden, R. G., Balogh, A. et al., 2003, "The Sun and Heliosphere at Solar Maximum", *Science*, **302**, 1165–1169. [DOI], [ADS] (Cited on page 33.)

Smith, S.M. and Schatten, K.H., 1970, "Preliminary Comparison of Predicted and Observed Structure of the Solar Corona at the Eclipse of 7 March 1970", *Bull. Am. Astron. Soc.*, **2**, 345. [ADS] (Cited on page 33.)

Solanki, S., Usoskin, I.G., Kromer, B., Schüssler, M. and Beer, J., 2005, "Reply", *Nature*, **436**, E4–E5. [DOI], [ADS] (Cited on page 59.)

Solanki, S.K., Schüssler, M. and Fligge, M., 2000, "Secular evolution of the Sun's magnetic field since the Maunder minimum", *Nature*, **480**, 445–446. [DOI], [ADS] (Cited on pages 57, 60, and 61.)

Solanki, S.K., Scüssler, M. and Fligge, M., 2002, "Secular evolution of the Sun's magnetic flux", *Astron. Astrophys.*, **383**, 706–712. [DOI], [ADS] (Cited on pages 60 and 61.)

Solanki, S.K., Usoskin, I.G., Kromer, B., Schüssler, M. and Beer, J., 2004, "Unusual activity of the Sun during recent decades compared to the previous 11,000 years", *Nature*, **431**, 1084–1087. [DOI], [ADS] (Cited on pages 57 and 59.)

Stamper, R., Lockwood, M., Wild, M.N. and Clark, T.D.G., 1999, "Solar causes of the long-term increase in geomagnetic activity", *J. Geophys. Res.*, **104**, 28,325–28,342. [DOI], [ADS] (Cited on pages 15 and 25.)

Steinhilber, F. and Beer, J., 2013, "Prediction of solar activity for the next 500 years", *J. Geophys. Res.*, **118**. [DOI] (Cited on pages 64 and 67.)

Steinhilber, F., Abreu, J.A. and Beer, J., 2008, "Solar modulation during the Holocene", *Astrophys. Space Sci. Trans.*, **4**, 1–6. [DOI], [ADS]. Online version (accessed 14 April 2013): http://hal.archives-ouvertes.fr/hal-00297488 (Cited on pages 58, 59, 63, 64, 65, and 66.)

Steinhilber, F., Abreu, J.A., Beer, J. and McCracken, K.G., 2010, "Interplanetary magnetic field during the past 9300 years inferred from cosmogenic radionuclides", *J. Geophys. Res.*, **115**, A01104. [DOI], [ADS]. Online version (accessed 14 April 2013): http://www.eawag.ch/forschung/surf/publikationen/2010/ (Cited on pages 42, 50, and 58.)

Stuiver, M. and Quay, P.D., 1980, "Changes in atmospheric carbon-14 attributed to a variable Sun", *Science*, **207**, 11–19. [DOI], [ADS] (Cited on page 57.)

Suess, S.T. and Smith, E.J., 1996, "Latitudinal dependence of the radial IMF component: Coronal imprint", *Geophys. Res. Lett.*, **23**, 3267–3270. [DOI], [ADS] (Cited on page 34.)

Suess, S.T., Smith, E.J., Phillips, J., Goldstein, B.E. and Nerney, S., 1996, "Latitudinal dependence of the radial IMF component: Interplanetary imprint", *Astron. Astrophys.*, **316**, 304–312. [ADS] (Cited on page 34.)

Suess, S.T., Phillips, J.L., McComas, D.J., Goldstein, B.E., Neugebauer, M. and Nerney, S., 1998, "The solar wind – inner heliosphere", *Space Sci. Rev.*, **83**, 75–86. [DOI], [ADS] (Cited on page 34.)

Sugiura, M. and Kamei, T., 1991, *Equatorial Dst index 1957-1986*, ISGI Publications Office, Saint-Maur-des-Fosses, France (Cited on page 7.)

Svalgaard, L. and Cliver, E.W., 2005, "The IDV index: Its derivation and use in inferring long-term variations of the interplanetary magnetic field strength", *J. Geophys. Res.*, **110**, A12103. [DOI], [ADS] (Cited on pages 10, 11, 15, 19, 42, 43, and 72.)

Svalgaard, L. and Cliver, E.W., 2006, "Reply to the comment by M. Lockwood et al. on 'The IDV index: Its derivation and use in inferring long-term variations of the interplanetary magnetic field"', *J. Geophys. Res.*, **111**, A09110. [DOI], [ADS] (Cited on pages 42, 43, and 44.)

Svalgaard, L. and Cliver, E.W., 2007a, "Interhourly variability index of geomagnetic activity and its use in deriving the long-term variation of solar wind speed", *J. Geophys. Res.*, **112**, A10111. [DOI], [ADS], [0706.0961] (Cited on pages 12, 15, 16, 30, 42, 47, and 52.)

Svalgaard, L. and Cliver, E.W., 2007b, "A Floor in the Solar Wind Magnetic Field", *Astrophys. J.*, **661**, L203–L206. [DOI], [ADS], [0706.0961] (Cited on page 52.)

Svalgaard, L. and Cliver, E.W., 2010, "Heliospheric magnetic field 1835–2009", *J. Geophys. Res.*, **115**, A09111. [DOI], [ADS] (Cited on pages 11, 14, 17, 19, 20, 42, 46, 47, 48, 52, and 56.)

Svalgaard, L., Duvall Jr, T.L. and Scherrer, P.H., 1978, "The strength of the Sun's polar fields", *Solar Phys.*, **58**, 225–239. [DOI], [ADS] (Cited on page 36.)

Svalgaard, L., Cliver, E.W. and Le Sager, P., 2003, "Determination of interplanetary magnetic field strength, solar wind speed and EUV irradiance, 1890–2003", in *Solar Variability as an Input to the Earth's Environment*, (Ed.) Wilson, A., ESA Special Publication, 535, pp. 15–23, European Space Agency, Nordwijk. [ADS] (Cited on pages 12, 15, 16, 30, and 47.)

Svalgaard, L., Cliver, E. W. and Le Sager, P., 2004, "IHV: a new long-term geomagnetic index", *Adv. Space Res.*, **34**(2), 436–439. [DOI], [ADS] (Cited on pages 15 and 16.)

Svalgaard, L., Cliver, E.W. and Kamide, Y, 2005, "Sunspot cycle 24: Smallest cycle in 100 years?", *Geophys. Res. Lett.*, **32**, L01104. [DOI], [ADS] (Cited on page 53.)

Taricco, C., Bhandari, N., Cane, D., Colombetti, P. and Verma, N., 2006, "Galactic cosmic ray flux decline and periodicities in the interplanetary space during the last 3 centuries revealed by ^{44}Ti in meteorites", *J. Geophys. Res.*, **111**, A08102. [DOI], [ADS] (Cited on page 57.)

Thomas, S., Owens, M.J. and Lockwood, M., 2013, "The 22-Year Hale Cycle in Cosmic Ray Flux – Evidence for Direct Heliospheric Modulation", *Solar Phys.*. [DOI], [ADS], [arXiv:1304.5191] (Cited on page 57.)

Tomita, S., Nosé, M., Iyemori, T. et al., 2011, "Magnetic local time dependence of geomagnetic disturbances contributing to the AU and AL indices", *Ann. Geophys.*, **29**, 673–678. [DOI], [ADS] (Cited on pages 7, 9, and 27.)

Uchida, Y., Altschuler, M.D. and Newkirk Jr, G., 1973, "Flare-Produced Coronal MHD-Fast-Mode Wavefronts and Moreton's Wave Phenomenon", *Solar Phys.*, **28**, 495–516. [DOI], [ADS] (Cited on page 33.)

Ulrich, R.K., 1992, "Analysis of magnetic fluxtubes on the solar surface from observations at Mt. Wilson of λ5250 and λ5233", in *Cool Stars, Stellar Systems, and the Sun: Seventh Cambridge Workshop*, Tuscon, Arizona, October 9 – 12, 1991, (Eds.) Giampapa, M.S., Bookbinder, J.A., ASP Conference Series, 26, pp. 265–267, Astronomical Society of the Pacific, San Francisco. [ADS] (Cited on page 36.)

Usoskin, I.G., 2013, "A History of Solar Activity over Millennia", *Living Rev. Solar Phys.*, **10**, lrsp-2013-1. [DOI], [ADS]. URL (accessed 14 April 2013):
http://www.livingreviews.org/lrsp-2013-1 (Cited on page 57.)

Usoskin, I.G., Mursula, K. and Kovaltsov, G.A., 2001, "Heliospheric modulation of cosmic rays and solar activity during the Maunder Minimum", *J. Geophys. Res.*, **106**, 16,039–16,046. [DOI], [ADS] (Cited on page 62.)

Usoskin, I.G., Mursula, K., Solanki, S.K., Schüssler, M. and Kovaltsov, G.A., 2002, "A physical reconstruction of cosmic ray intensity since 1610", *J. Geophys. Res.*, **107**(A11), 1374. [DOI], [ADS]. Online version (accessed 14 April 2013):
http://cc.oulu.fi/~usoskin/personal/ (Cited on pages 65 and 66.)

Usoskin, I.G., Solanki, S.K., Schüssler, M., Mursula, K. and Alanko, K., 2003, "Millennium-Scale Sunspot Number Reconstruction: Evidence for an Unusually Active Sun since the 1940s", *Phys. Rev. Lett.*, **91**, 211101. [DOI], [ADS], [astro-ph/0310823] (Cited on page 59.)

Usoskin, I.G., Mursula, K., Solanki, S.K., Schüssler, M. and Alanko, K., 2004, "Reconstruction of solar activity for the last millennium using 10Be data", *Astron. Astrophys.*, **413**, 745–751. [DOI], [ADS], [astro-ph/0309556] (Cited on page 59.)

Usoskin, I.G., Alanko-Huotari, K., Kovaltsov, G.A. and Mursula, K., 2005, "Heliospheric modulation of cosmic rays: Monthly reconstruction for 1951-2004", *J. Geophys. Res.*, **110**, A12 108. [DOI], [ADS] (Cited on page 57.)

Usoskin, I.G., Solanki, S.K., Taricco, C., Bhandari, N. and Kovaltsov, G.A., 2006, "Long-term solar activity reconstructions: direct test by cosmogenic ^{44}Ti in meteorites", *Astron. Astrophys.*, **457**, L25–L28. [DOI], [ADS] (Cited on page 57.)

Vaquero, J.M., Gallego, M.C., Usoskin, I.G. and Kovaltsov, G.A., 2011, "Revisited sunspot data: A new scenario for the onset of the Maunder minimum", *Astrophys. J.*, **731**, L24. [DOI], [ADS] (Cited on page 51.)

Vasyluinas, V.M., Kan, J.R., Siscoe, G.L. and Akasofu, S.-I., 1982, "Scaling relations governing magnetospheric energy transfer", *Planet. Space Sci.*, **30**, 359–365. [DOI], [ADS] (Cited on page 22.)

Vieira, L.E.A. and Solanki, S.K., 2010, "Evolution of the solar magnetic flux on time scales of years to millenia", *Astron. Astrophys.*, **509**, A100. [DOI], [ADS] (Cited on page 61.)

Vogel, R.L., 1998, "Why scientists have not been writing metadata", *Eos Trans. AGU*, **79**(31), 373–380. [DOI], [ADS] (Cited on page 21.)

Wang, Y.-M. and Sheeley Jr, N.R., 1990, "Magnetic flux transport and the sunspot-cycle evolution of coronal holes and their wind streams", *Astrophys. J.*, **365**, 372–386. [DOI], [ADS] (Cited on pages 33 and 46.)

Wang, Y.-M. and Sheeley Jr, N.R., 1995, "Solar implications of Ulysses interplanetary field measurements", *Astrophys. J.*, **447**, L143–L146. [DOI], [ADS] (Cited on pages 36 and 37.)

Wang, Y.-M. and Sheeley Jr, N.R., 2002, "Sunspot activity and the long-term variation of the Sun's open magnetic flux", *J. Geophys. Res.*, **107**, 1302. [DOI], [ADS] (Cited on page 60.)

Wang, Y.-M. and Sheeley Jr, N.R., 2003a, "On the fluctuating component of the sun's large-scale magnetic field", *Astrophys. J.*, **590**, 1111–1120. [DOI], [ADS] (Cited on pages 36 and 37.)

Wang, Y.-M. and Sheeley Jr, N.R., 2003b, "On the Topological Evolution of the Coronal Magnetic Field During the Solar Cycle", *Astrophys. J.*, **599**, 1404–1417. [DOI], [ADS] (Cited on page 60.)

Wang, Y.-M. and Sheeley Jr, N.R., 2012, "The solar wind and interplanetary field during very low amplitude sunspot cycles", *Astrophys. J.*, **764**, 90. [DOI], [ADS] (Cited on page 47.)

Wang, Y.-M., Sheeley Jr, N.R., Howard, R.A and Rich, N.B., 1999a, "Streamer disconnection events observed with the LASCO coronograph", *Geophys. Res. Lett.*, **26**, 1349–1352. [DOI], [ADS] (Cited on page 61.)

Wang, Y.-M., Sheeley Jr, N.R., Howard, R.A. and St Cyr, O.C., 1999b, "Coronograph observations of inflows during high solar activity", *Geophys. Res. Lett.*, **26**, 1203–1206. [DOI], [ADS] (Cited on page 61.)

Wang, Y.-M., Lean, J. and Sheeley Jr, N.R., 2000, "The long-term evolution of the Sun's open magnetic flux", *Geophys. Res. Lett.*, **27**, 505–508. [DOI], [ADS] (Cited on page 36.)

Wang, Y.-M., Lean, J. and Sheeley Jr, N.R., 2002, "Role of a variable meridional flow in the secular evolution of the sun's polar fields and open flux", *Astrophys. J.*, **577**, L53–L57. [DOI], [ADS] (Cited on page 60.)

Wang, Y.-M., Lean, J. and Sheeley Jr, N.R., 2005, "Modeling the Sun's Magnetic Field and Irradiance since 1713", *Astrophys. J.*, **625**(1), 522–538. [DOI], [ADS] (Cited on page 60.)

Zhou, X.-Y., Tsurutani, B.T., Reeves, G. et al., 2003, "Ring current intensification and convection-driven negative bays: Multisatellite studies", *J. Geophys. Res.*, **108**, 1407. [DOI], [ADS] (Cited on page 27.)