



RESEARCH ARTICLE

10.1002/2014JA019970

This article is a companion to *Lockwood et al.* [2014] doi:10.1002/2014JA019972; *Lockwood and Owens* [2014] doi:10.1002/2014JA019973.

Key Points:

- Correction of the Zürich sunspot number record around 1945 is required
- The best estimate of the correction factor required is 12%
- The probability that the correction required is as large as 20% is 0.00001

Correspondence to:

M. Lockwood,
m.lockwood@reading.ac.uk

Citation:

Lockwood, M., M. J. Owens, and L. Barnard (2014), Centennial variations in sunspot number, open solar flux, and streamer belt width: 1. Correction of the sunspot number record since 1874, *J. Geophys. Res. Space Physics*, 119, doi:10.1002/2014JA019970.

Received 12 MAR 2014

Accepted 10 JUN 2014

Accepted article online 11 JUN 2014

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Centennial variations in sunspot number, open solar flux, and streamer belt width: 1. Correction of the sunspot number record since 1874

M. Lockwood¹, M. J. Owens¹, and L. Barnard¹

¹Department of Meteorology, University of Reading, Reading, UK

Abstract We analyze the widely used international/Zürich sunspot number record, R , with a view to quantifying a suspected calibration discontinuity around 1945 (which has been termed the “Waldmeier discontinuity”). We compare R against the composite sunspot group data from the Royal Greenwich Observatory network and the Solar Optical Observing Network, using both the number of sunspot groups, N_G , and the total area of the sunspots, A_G . In addition, we compare R with the recently developed interdiurnal variability geomagnetic indices IDV and $IDV(1d)$. In all four cases, linearity of the relationship with R is not assumed and care is taken to ensure that the relationship of each with R is the same before and after the putative calibration change. It is shown the probability that a correction is not needed is of order 10^{-8} and that R is indeed too low before 1945. The optimum correction to R for values before 1945 is found to be 11.6%, 11.7%, 10.3%, and 7.9% using A_G , N_G , IDV , and $IDV(1d)$, respectively. The optimum value obtained by combining the sunspot group data is 11.6% with an uncertainty range 8.1–14.8% at the 2σ level. The geomagnetic indices provide an independent yet less stringent test but do give values that fall within the 2σ uncertainty band with optimum values are slightly lower than from the sunspot group data. The probability of the correction needed being as large as 20%, as advocated by Svalgaard (2011), is shown to be 1.6×10^{-5} .

1. Introduction

This is the first of a series of three papers to model the long-term variation in open solar flux and the variation in the latitudinal width of the coronal streamer belt. The model used in paper 3 [Lockwood and Owens, 2014] is a development of that by Owens and Lockwood [2012], which is a variant of the application of the open flux continuity equation by Solanki et al. [2000]. The development of open solar flux reconstructions and their modeling has recently been reviewed by Lockwood [2013] and Owens and Forsyth [2013] and the modeling in paper 3 reproduces the reconstructions by Lockwood et al. [2014b]. As in the Solanki et al. paper, the open flux emergence rate is modeled using sunspot number, although the precise quantification is different, being based on average flux content and occurrence frequency of coronal mass ejections. Thus, it is important that the input sunspot record is as accurate as possible. In this first paper, we discuss in detail the calibration of international/Zürich sunspot number across a putative calibration error around 1945.

A survey of sunspot metrics and catalogues has recently been presented by Clette et al. [2007] and Lefevre and Clette [2014]. The international sunspot number, R , quantifies the sunspot activity present on the surface of the Sun according to the formula

$$R = k(10N_G + N_S) \quad (1)$$

where N_G is the number of sunspot groups, N_S is the number of individual spots, and k is the observatory factor that varies with location and observing instrument. This has been compiled since 1981 from a number of observers (currently it uses 86 observers located in 29 different countries) and is designed to be a continuation of the Zürich sunspot numbers which were based upon observations made at Zürich and its two branch stations in Arosa and Locarno. The basis of this formulation in equation (1) has been in use since it was first introduced in 1848 by Rudolf Wolf. The Wolf and Zürich sunspot numbers used just one station (and, by definition, the initial Wolf values used $k = 1$), whereas the International sunspot number uses many observers, each with their own k value. The algorithm has also been applied retrospectively back to 1749. Daily values of an intercalibrated composite of Wolf, Zürich, and international sunspot numbers are available for 1818 onward from the World Data Centre for the Sunspot Index and Long-term Solar Observations (WDC-SILSO) of the Royal Observatory of Belgium,

Brussels. Before this date, daily observations are too sparse but monthly values have been compiled. In the present paper we refer to the intercalibrated WDC-SILSO composite as R . The need to scale N_G and N_S from observations, sketches, or photographic plates leads to some degree of subjectivity, but this is limited in the modern data: the RMS dispersion between simultaneous daily observations by many separate observers, including the effect of different meteorological conditions, is of order 8%. Nevertheless, subjective decisions about what constitutes a group must be made and a drift in the k factor could lead to a calibration error. *Clette et al.* [2007] have reviewed the history of the compilation of R and discussed potential sources of error. In this first paper we consider the accuracy of R after 1874, when photographic observations by the Royal Greenwich Observatory commenced. The WDC-SILSO composite in this interval uses international sunspot number for 1981 to the present, Zürich sunspot number for 1882–1980, and Wolf sunspot numbers for 1874–1882 (which were intercalibrated with Zürich numbers over the interval 1877–1892 by A. Wolf (see Paper 2 [Lockwood et al., 2014c]). In Paper 2 we study sunspot numbers before 1874.

There are several complications with the sunspot number record. The early Zürich/Wolf sunspot number shows well-known long-term differences to the group sunspot number R_G defined by *Hoyt and Schatten* [1994, 1998] as

$$R_G = \langle k' \times 12.08 \times N_G \rangle_N \quad (2)$$

where N_G is number of sunspot groups and k' is a site/observer factor and the averaging is done over the N observers that are available for that day. Daily values are then averaged to give annual means. The factor 12.08 was derived to make R_G and modern international sunspot number R the same on average. Because of the greater difficulty in quantifying the number of individual spots, N_S , the group number is generally regarded as more robust than the early Wolf number (see discussion by, for example, *Hathaway* [2010] and *Usoskin* [2013]). However, because, in general, N_S is not necessarily proportional to N_G on all time scales, it should be noted that R_G and R are not identical metrics of solar activity. Hence, differences can be real and do not necessarily indicate errors in one or both data sequences. *Leussu et al.* [2013] have used Schwabe's original data to identify an inconsistency in the Wolf sunspot number around 1849, associated with the change of the primary observer from Schwabe to Wolf. This calls for a reduction in R before 1848 by 20% which accounts for much of the large difference between R_G and R before this date.

Recently, *Svalgaard* [2011] has argued that the international sunspot number record in common usage, and as stored in many data centers, suffers from another calibration problem around 1945 with values before then being 20% too low. *Svalgaard* terms this the “Waldmeier discontinuity” [see also *Aparicio et al.* 2012 and *Cliver et al.* 2013] and argues that it was caused by the introduction of a weighting scheme for sunspot counts according to their size and a change in the procedure used to define a group; both changes that may have been introduced by the then director of the Zürich observatory, Max Waldmeier, when he took over responsibility for the production of the Wolf sunspot number in 1945. *Svalgaard* argues that these corrections were not applied before this date, despite Waldmeier's claims to the contrary. By comparison with other long time series of solar and solar-terrestrial indices, *Svalgaard* makes a compelling case that this discontinuity is indeed present in the data. He argues that sunspot number values before 1945 need to be increased by 20% as a result.

One issue that needs careful consideration in quantifying the correction needed is the homogeneity of the data series that R is compared to, both before and after the putative discontinuity. Also, it is very important that the indices compared are constructed completely independent of R , if they are to act as a test and/or calibration of R . Because sunspot number has been used as a standard for many years, this is more of a problem than it first appears to be. For example, there is evidence that sunspot data were used to correct geomagnetic observations and such cross contamination is highly undesirable.

In this paper, we use four indices for which the provenance is well known. The time series of annual means of these data are shown in Figure 1. Two of these (specifically the number of sunspot groups, N_G , and the total area of sunspots, A_G) are taken from the sunspot group data generated from 1874 onward by the Royal Greenwich Observatory (RGO) based on photographic observations at Greenwich and several support stations (Cape of Good Hope, Kodaikanal, and Mauritius), used to minimize data loss due to cloud cover. Differences between data sets can arise due to errors introduced by the personal bias of the observer in defining a group, limited seeing conditions at the observation site, different amounts of scattered light, the difference in the time when the observations were made, and the instrumentation and techniques deployed.

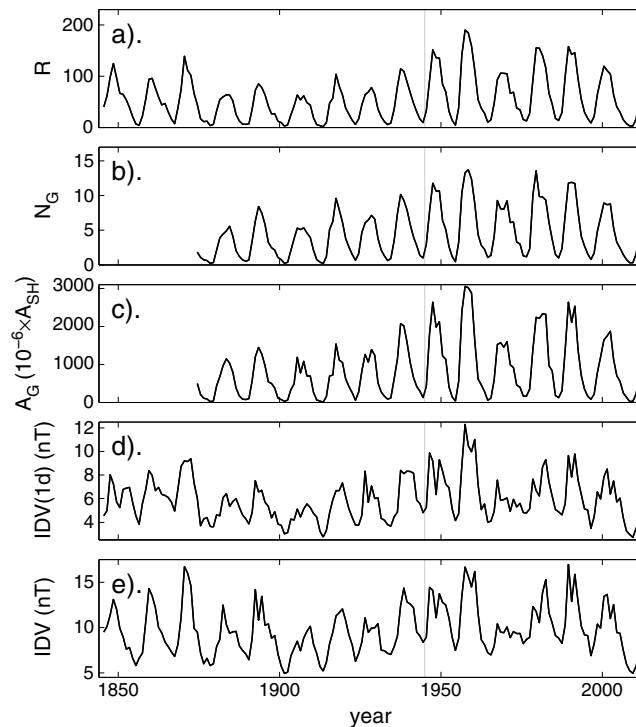


Figure 1. Time variations of annual means for 1845 to 2012 (inclusive). (a) The international sunspot number, R . (b) The daily number of sunspot groups N_G from the RGO/SOON data set. (c) The total daily sunspot group area A_G (corrected for the foreshortening effect of position on the solar disk and given in millionths of the area of a solar hemisphere, A_{SH}) from the RGO/SOON data set. Both the N_G and A_G series use the intercalibration between the RGO and SOON data derived in section 2 of this paper. (d) The $IDV(1d)$ geomagnetic index generated by Lockwood *et al.* [2013], with the correction of Lockwood *et al.* [2014a] to solar cycle 11 applied. (e) The IDV geomagnetic index of Svalgaard and Cliver [2010]. The vertical grey lines are at 1 January 1946.

[2001] have shown this leads to systematically lower A_G values. Foster [2004] and Balmaceda *et al.* [2009] used the Pulkovo Observatory SD data, along with some other data sources, to achieve intercalibration between the RGO and SOON sunspot areas. In section 2 of this paper, we also consider intercalibration of the RGO and SOON data. In section 3 we study the 1945 discontinuity using both the combined RGO/SOON data and the RGO data alone.

The other data used in the present paper are interdiurnal variation geomagnetic indices. Svalgaard and Cliver [2010] compiled the IDV index based on Bartels' u index [Bartels, 1932]. Lockwood *et al.* [2013] have generated a variant of this index, $IDV(1d)$ designed to be as homogeneous in its construction as possible. In this paper, we make use of both indices after 1874, when they are extremely similar despite considerable differences in how they are compiled and the different data sets used. We chose these indices because both depend primarily on the strength of the near-Earth interplanetary magnetic field (IMF) without the strong dependence on solar wind speed found in range indices such as aa and Ap (see review by Lockwood [2013]): this means that they also have a simpler relationship to sunspot number. Lockwood *et al.* [2014a] show that modeled and observed IMF data series correlate with N_G^n with optimum n of 0.4 and 0.5 for observed and geomagnetically reconstructed IMF, respectively. In both cases, the simultaneous N_G^n value explains about 70% of the IMF variation in annual means, the remainder being mainly associated with the prior history of N_G . In section 4 we employ the relationship between IDV and $IDV(1d)$ and the simultaneous N_G value to provide two further tests of R . Further aspects of the relationship between sunspot number and the interdiurnal geomagnetic indices are discussed in Paper 2 of this series [Lockwood *et al.*, 2014c].

These errors can be both random and systematic. Because the original photographic plates still exist, the derived sunspot data sets can be checked and various typographic, systematic, and isolated errors have already been identified [Willis *et al.*, 2013] and a list for before 1917 has been given by Erwin *et al.* [2013]. Many corrections relate to the locations of spot groups on the visible disk rather than their area and number, and at the time of writing a corrected digital data are not yet available. However, Erwin *et al.* [2013] report that errors in existing data sets are "remarkably small."

Unfortunately, the UK RGO observation series was terminated in 1976. Subsequently, the observation series has been taken on by the Solar Optical Observing Network (SOON), funded by the United States Air Force and NOAA with some data gaps that are filled using the "Solnechniye Danniyе" (Solar Data, SD) Bulletins issued by the Pulkovo Astronomical Observatory in Russia. The main SOON observations are from Boulder with support stations at Holloman, Learmonth, Palehua, Ramey, San Vito, and Mount Wilson. Unlike the RGO data, most SOON data are available as drawings rather than photographic plates and Baranyi *et al.*

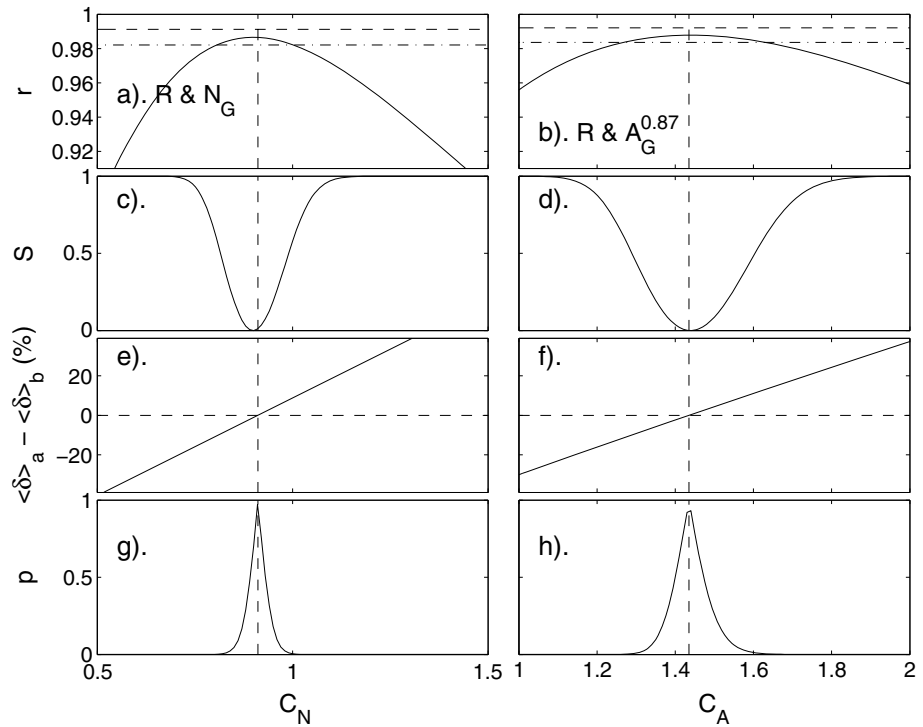


Figure 2. Analysis of the intercalibration of the RGO and SOON sunspot group data using data for 1946–2012. (a, c, e, and g) The number of groups, N_G . (b, d, f, and h) The total sunspot area, A_G^m for the best fit exponent m of 0.87. All panels are plotted as a function of the gain factor C by which the SOON data need multiplication, which was varied between 0.5 and 2. Figures 2a and 2b show the correlation coefficients of the full composite (1946–2012) with R , with the horizontal dashed line showing the correlation for the RGO data and the dash-dotted line showing that for the SOON data. Figures 2c and 2d show the significance S of the difference between the peak correlation and that at general C . Figures 2e and 2f show the difference between the mean residual of the best fit to R for the SOON data, $\langle \delta \rangle_a$, and that for the RGO data, $\langle \delta \rangle_b$, shown as a percentage of the overall mean (for $C = 1$) of N_G in Figure 2e and of A_G^m in Figure 2f. Figures 2g and 2h show the p value that $\langle \delta \rangle_b$ and $\langle \delta \rangle_a$ are the same, computed using Welch’s t test. The vertical dashed lines mark the peak p values in both cases.

The grey vertical line in each panel of Figure 1 marks the date of the putative Waldmeier discontinuity. There is no obvious major change in the behavior of the sunspot number, the sunspot group data, or the geomagnetic indices around this time; however, it lies on the steep initial rising phase of a solar cycle when a change in a scaling factor would be hardest to detect.

2. Intercalibration of the RGO and SOON Data Sets

Because the focus of the present paper is the sunspot number R , we use this to intercalibrate the RGO and SOON data, rather than a series of intercalibrations of overlapping data. This method was actually employed in some places by *Balmaceda et al.* [2009], in addition to their use of overlapping data series. Figure 2 demonstrates the two methods used to search for calibration changes that are employed in this paper. In this section, both methods are used to investigate the intercalibration of the RGO and SOON sunspot group data. We employ only data from after the putative Waldmeier discontinuity in R around 1945 to intercalibrate the RGO and SOON data.

Figures 2a, 2c, 2e, and 2g are for the number of groups, N_G , and Figures 2b, 2d, 2f, and 2h are for the total sunspot area, A_G^m , for the best fit exponent m of 0.87 (derived below). All panels are plotted as a function of the gain factors C by which the SOON data need multiplication, such that $N_G[\text{RGO}] \equiv C_N \times N_G[\text{SOON}]$ and $A_G[\text{RGO}]^m \equiv C_A \times A_G[\text{SOON}]^m$. Both C_A and C_N were varied between 0.5 and 2.

2.1. Analysis Using Correlation

Figures 2a and 2b show the variation of the correlation coefficients r of the full RGO/SOON composite with R for the interval 1946–2012 (i.e., after the putative discontinuity in R) with correction coefficient that has

been applied to the SOON data (C_N in the case of N_G and C_A in the case of A_G^m). The horizontal dashed lines show the correlation for the RGO data alone with the horizontal dash-dotted lines showing that for the SOON data alone. These correlograms both show clear peaks, but the question arises how well this defines the optimum values of C_N and C_A , because it is not immediately apparent at what value each correlation is significantly lower than its peak value. This question is addressed in Figures 2c and 2d which show the significance S of the difference between the peak correlation and that at general C , computed using the Fisher Z test and assessed by comparison with the AR-1 autoregressive red noise model using the procedure of Lockwood [2002]. It can be seen the peak correlation lies between those for the two data series individually (which is slightly higher for the RGO data in both cases). However, as C is varied away from the optimum values, r falls because of the discontinuity introduced between the data sets by the nonoptimum values of C . S is a measure of the statistical significance of that decline.

2.2. Analysis Using Fit Residuals

Figures 2e and 2f study the mean residuals of the best fit to R using the full data sequences of N_G or A_G^m (for 1946–2012). These plots show the differences $\langle \delta \rangle_a - \langle \delta \rangle_b$ where $\langle \delta \rangle_a$ is the mean residual for the SOON data (1977–2012) and $\langle \delta \rangle_b$ is the mean residual for the RGO data used (1946–1976). Figure 2e is for N_G , and Figure 2f is for A_G^m . To normalize these differences, they are shown as a percentage of the overall mean for $C = 1$. For the optimum C , the difference should fall to zero, i.e., there is no change in mean residual across the RGO/SOON data join. These values are marked by the vertical dashed lines, being at $C = C_N = 0.911$ for N_G and $C = C_A = 1.435$ for A_G ; these values are very close indeed to the C values giving peak correlations in Figures 2a and 2b (Figures 2c and 2d show that the significance of any difference is very close indeed to zero). Figures 2g and 2h show the probability p value that $\langle \delta \rangle_b$ and $\langle \delta \rangle_a$ are the same. Because both the sample sizes and the variances are not the same for the two data subsets, we use Welch's t test to evaluate the p values of the difference between the mean fit residual before and after 1 January 1946; this two-sample t test is a parametric test that compares two independent data samples [Welch, 1947]. Because it is not assumed that the two data samples are from populations with equal variances, the test statistic under the null hypothesis has an approximate Student's t distribution with a number of degrees of freedom given by Satterthwaite's approximation [Satterthwaite, 1946]. The distributions of residuals are close to Gaussian distributions in this case and so application of nonparametric tests (specifically, the Mann-Whitney U (Wilcoxon) test of the medians and the Kolmogorov-Smirnov test of the overall distributions) gives very similar results (not shown).

Using overlap of the Pulkovo data, *Balmaceda et al.* [2009] obtained a value of $C_A = 1.489 \pm 0.194$ (for A_G^m , with $m = 0.732$) and the value of 1.435 derived here lies comfortably within their uncertainty band. The main difference in the optimum value arises because we have used R to intercalibrate whereas *Balmaceda et al.* used, in the main, overlapping data. In addition, we have only used data for after the putative Waldmeier discontinuity in 1945. The great consistency between the two methods demonstrates the great stability of R during the interval studied (1946–2012). *Hathaway* [2010] derives and uses a value of 1.48.

Figure 3 shows scatterplots of annual means of R as a function of (a) N_G and (b) A_G for all data from after the putative Waldmeier. The solid points are the RGO data, and the open circles are the SOON data for correction factors $C_N = 0.911$ for N_G and $C_A = 1.435$ for A_G^m . It can be seen that the behavior for both RGO and corrected SOON data is the same after the putative Waldmeier discontinuity, with an almost linear variation for N_G (but slightly greater scatter) and a somewhat less linear variation for A_G .

3. Relationships Between R and Calibration Time Series

3.1. Sunspot Group Data

Figure 4 studies the relationships between R and N_G and between R and A_G before, after, and across the putative discontinuity in 1945. For both N_G and A_G , the optimum RGO/SOON calibration factors (C_N and C_A , respectively) derived in section 2 are applied.

In Figure 4, several parameters are shown as a function of an exponent m , which is varied between 0 and 1.5. Figures 4a and 4b show the correlation between R and (a) N_G^m and (b) A_G^m . The exponent m is required as neither plot in Figure 3 is self-evidently fully linear. The red lines are for before the putative discontinuity (1874–1945, inclusive), the blue lines are for after it (1946–2012, inclusive), and the black lines are for the

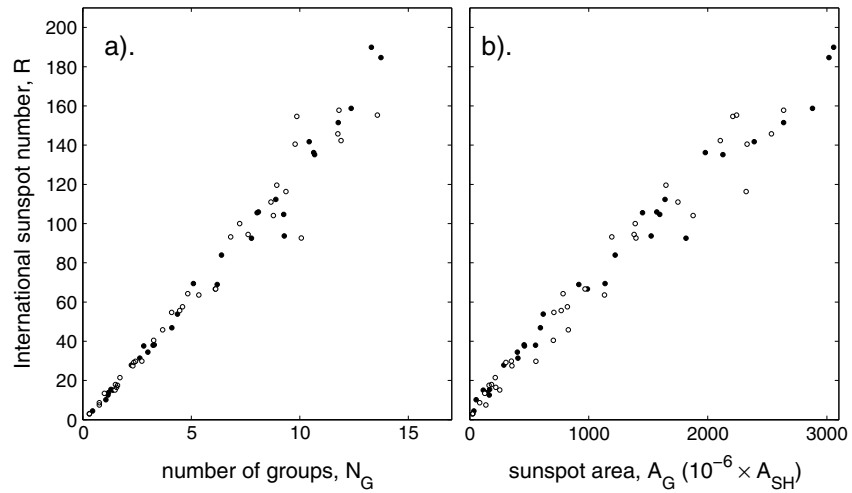


Figure 3. Scatterplots of annual means of the International sunspot number R , as a function of the RGO/SOON sunspot group observations for 1946–2012 (so only data for after the putative Waldmeier discontinuity are shown): (a) the number of sunspot groups, N_G ; (b) the total sunspot area (corrected for foreshortening) A_G (in units of millionths of the surface area of a solar hemisphere, A_{SH}). Data from RGO (1946–1976) are shown as solid points, data from the SOON network (1977–2012) as open circles. The optimum values of C_N and C_A derived in Figure 2 have been applied to the SOON data.

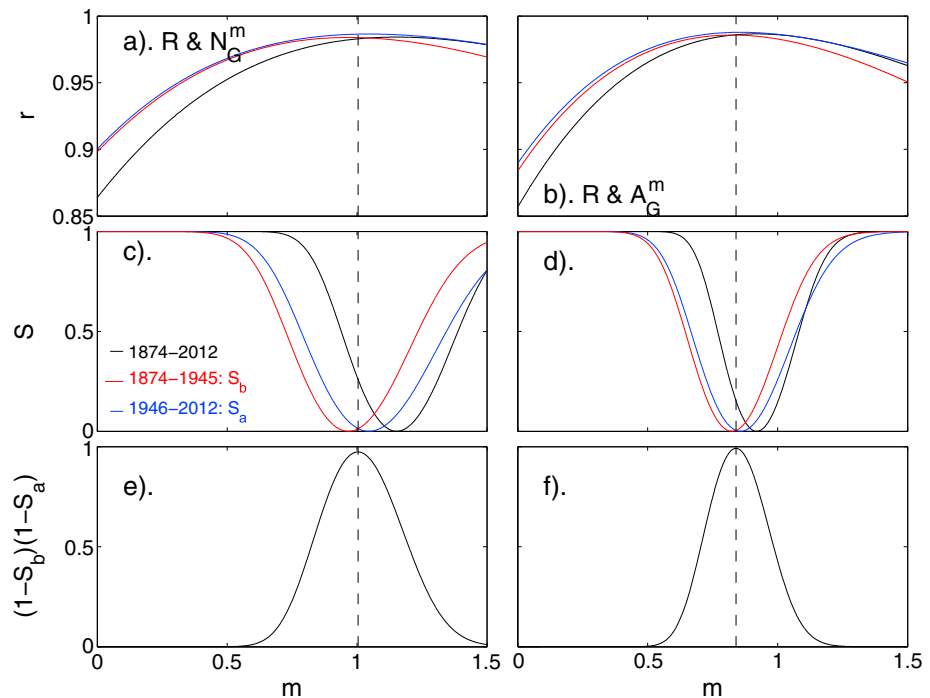


Figure 4. Analysis of the relationships between annual means of the RGO/SOON sunspot group number, N_G , and total area, A_G , and international sunspot number, R . (a, c, and e) For N_G . (b, d, and f) For A_G . In Figures 3a–3d, black lines are for all data (1874–2012), red lines are for before 1946, and blue lines are for after (and including) 1946. Figure 3a shows the correlation coefficient between R and N_G^m and Figure 3b the correlation coefficient between R and A_G^m , both as a function of the exponent m . Figures 3c and 3d are the significance S of the difference between the peak correlation and that at other m (S_b is for before 1946; S_a is for after that date), evaluated using the Fisher Z test. Figures 3e and 3f are the probability that the data for before and after 1 January 1946 have, in reality, the same value of m , $(1 - S_b)(1 - S_a)$, as a function of m . The vertical dashed lines mark the peaks in Figures 3e and 3f and are at $m = 1.00$ and $m = 0.87$ for R_G and A_G , respectively.

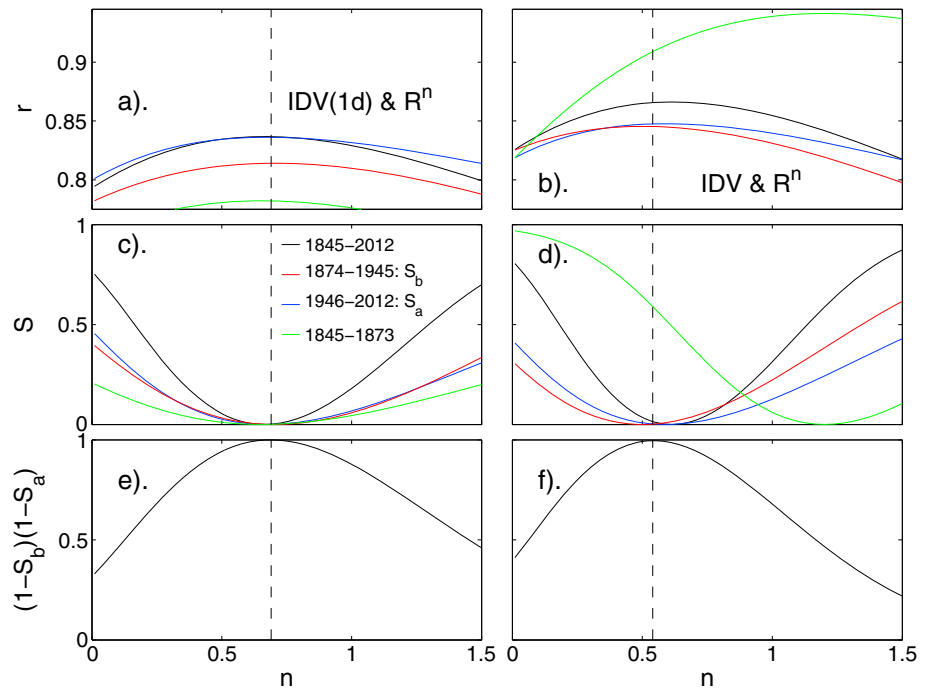


Figure 5. Analysis of the relationship between interdiurnal variation geomagnetic indices, IDV and $IDV(1d)$ with R^n for a range of exponents n . (a, c, and e) For $IDV(1d)$. (b, d, and f) For IDV . In Figures 5a–5d, black lines are for all data (1845–2012), green lines are for 1845–1874, red lines are for 1874–1945, and blue lines are for 1946–2012. Figures 5a and 5b are correlation coefficients r . Figures 5c and 5d are the significance S of the difference between the peak correlation and that at other n (S_b is for 1874–1945; S_a is for 1946–2012). Figures 5e and 5f are the probability that the data for before and after 1 January 1946 have, in reality, the same value of n , $(1 - S_b)(1 - S_a)$. The vertical dashed lines mark the peaks in Figures 5e and 5f and are at $n=0.54$ and $n=0.69$ for $IDV(1d)$ and IDV , respectively.

whole data set (1874–2012). In the case of A_G^m (Figure 4b), the behavior before and after 1 January 1946 is almost identical with peak correlation at $m = 0.87$ in both data subsets. However, the behavior of the total data set (1874–2012, black line) is different with the peak correlation at a slightly larger m . This is a clear indication that there is a discontinuity in either R or A_G (or both) around 1945–1946. In the case of N_G^m (Figure 4a), the behavior of the before and after subsets is not as similar as in the case of A_G , with the before data set giving slightly lower correlations that peak at a lower m . As expected from the scatter in Figure 3, peak correlations are slightly lower for N_G than A_G . Again, the correlation for the total data set is different from that of the two subsets individually, again showing a discontinuity in one or both data series around 1 January 1946. Figures 4c and 4d show the significances S of the difference between the three peak correlations and their values at general m , generated in the same way as Figures 2c and 2d, where S_b is for the data from before the putative discontinuity and S_a is for the data from after it. Figures 4e and 4f are used to derive the optimum values of m , by showing the probability that the data for before and after share the same m , $(1 - S_a)(1 - S_b)$ as a function of m . The peak of this probability is essentially unity for A_G and gives the optimum m of 0.87. For N_G the peak is very slightly lower at 0.96 and is at $m = 1.00$.

3.2. Geomagnetic Data

Figure 5 repeats this process using the geomagnetic indices IDV and $IDV(1d)$. In this case, the behavior is more clearly seen if we correlate these geomagnetic indices with R^n because $n < 1$ for all best fits. We here add a third time period, 1845–1873, for which both R and geomagnetic data are available, results for which are shown in green. For $IDV(1d)$ the behavior is the same in all three intervals with peak correlation at $n = 0.54$: correlations get progressively lower for the earlier intervals, which is to be expected as measurement noise should have decreased over time for both the sunspot and geomagnetic data series as observing techniques improved. For IDV , Figure 5b shows that the 1874–1945 and 1946–2012 intervals again give very similar behavior, but that for 1845–1873 is very different indeed (green line). In particular, this interval gives a much larger correlation coefficient and at an n value near unity. This demonstrates that IDV for this interval is much

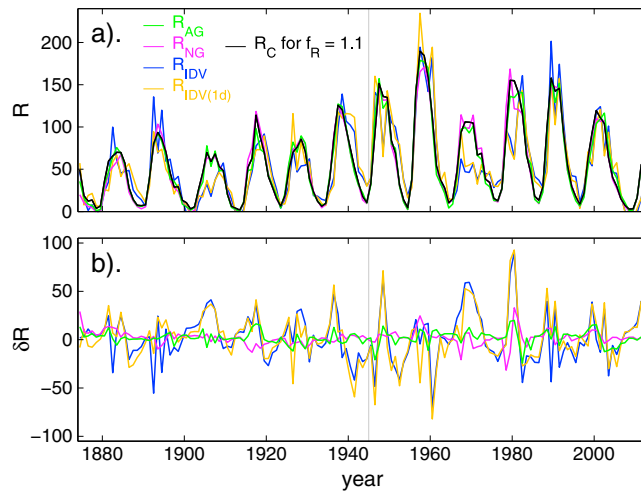


Figure 6. (a) Best fit variations of sunspot number for the example of $f_R = 1.100$. The black line is the corrected international sunspot number, R_C , for $f_R = 1.100$ (i.e., $R_C = R$ for 1946 and after and $R_C = 1.1R$ for 1945 and before). The best fit variations are shown from (green) A_G , (mauve) N_G , (blue) IDV , and (orange) $IDV(1d)$. (b) The fit residuals, δR (R_C minus the best fit variations), shown using the same colors as in Figure 6a. The vertical grey line is at 1 January 1946.

more similar to R than for later intervals. In fact, IDV from this period is based on a different type of data than for later periods, being $10u$, where u is Bartel's u index which, for much of the interval in question, was compiled from the diurnal range of geomagnetic activity [see *Svalgaard and Cliver, 2010*]; hence, IDV is not a homogeneously constructed index, whereas *Lockwood et al. [2013]* devised $IDV(1d)$ to be as homogeneous in its construction as it possibly could be. In this paper, we use only data after 1874 to avoid these problems and to match the interval used for the sunspot group data, but the implications of Figure 5 will be discussed again in Paper 2 [*Lockwood et al., 2014c*]. Like $IDV(1d)$, IDV shows the same behavior with R in the 1874–1945 and 1946–2012 intervals and using the probability distribution that they have the same exponent n given in Figure 5f, we find the optimum n is 0.69 for IDV .

The relationship between R and the geomagnetic indices is actually more complex than it appears here because of the role of past history of solar activity [*Lockwood et al., 2014a*]. The time scale of the history that matters is of interest: The original modeling of open solar flux by *Solanki et al. [2000]* derived a loss time constant of 4 years, and similar modeling by *Lockwood [2003]* gave 3.6 years. In Paper 3 of this series [*Lockwood and Owens, 2014*], we derive a loss time constant that varies over the solar cycle between 0.7 years and 1.3 years giving a somewhat smaller average value. *Lockwood et al. [2011]* show that the resulting open solar flux predictability falls from 1 to 0.5 over 22 years, and *Lockwood et al. [2014a]* detect an influence on the near-Earth IMF of the mean sunspot number over the previous 11 years as well as the current value. We here use the current value only to evaluate average levels of sunspot number before and after the putative 1945 discontinuity: in the next section we show that the results are consistent with the findings from the sunspot group data, but the uncertainties are much greater because of the additional history factor.

4. Analysis of the Amplitude of the Putative Discontinuity in International Sunspot Number

The values of the exponents n and m derived in section 3 are here used to analyze the amplitude of the putative discontinuity around 1945 in the R record. The procedure used is this: the sunspot number is amended before 1945 by multiplying by a correction factor f_R which is varied between 0.9 and 1.3 (i.e., a change between a 10% decrease and a 30% increase is implemented). For each f_R this generates a sequence R_C , where $R_C = R$ for 1946 and after and $R_C = f_R \times R$ for 1945 and before. R_C is then linearly regressed with A_G^m , N_G^m , $IDV^{1/n}$, and $IDV(1d)^{1/n}$, using the relevant values of m and n derived in the last section ($m = 0.871$ for A_G , $m = 1.000$ for N_G , $n = 0.540$ for $IDV(1d)$, and $n = 0.690$ for IDV).

The fit residuals δ are then evaluated. For example, for A_G , the residuals are given by

$$\delta = R_C - R_{AG} = R_C - (s A_G^m + c) \tag{3}$$

where s and c are the best fit linear regression coefficients. The mean value of δ for before the discontinuity (1847–1945), $\langle \delta \rangle_b$, is then compared to that for after the discontinuity (1946–2012), $\langle \delta \rangle_a$, and the difference between the two quantified and its significance evaluated using Welch's t test, as discussed above. This procedure is applied to A_G , N_G , IDV , and $IDV(1d)$. An example of the results is presented in Figure 6, which is for $f_R = 1.100$. Figure 6a shows the time variations of R_C along with best fit regressions which are termed R_{AG} , R_{NG} , R_{IDV} , and $R_{IDV(1d)}$. Figure 6b shows the time series of the fit residuals δ for each case.

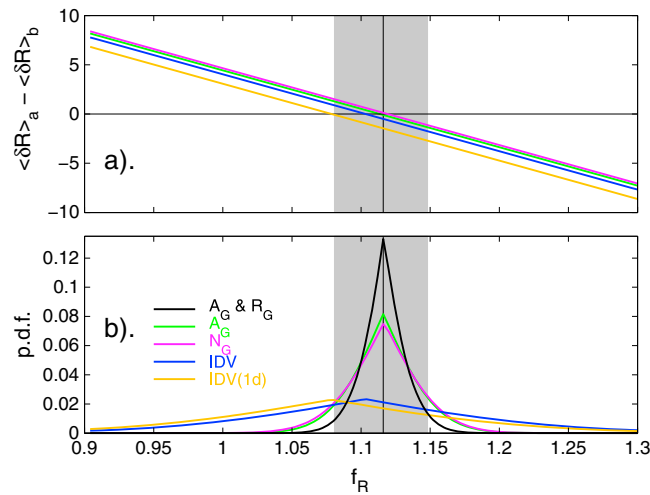


Figure 7. (a) The difference between the mean fit residuals after 1 January 1946 $\langle \delta R \rangle_a$ and before that date $\langle \delta R \rangle_b$ as a function of the f_R value used to generate R_C from R . The values of $\langle \delta R \rangle_a - \langle \delta R \rangle_b$ are shown for the best fits from (green) A_G , (mauve) N_G , (blue) IDV , and (orange) $IDV(1d)$. The optimum f_R value in each case is where the line crosses the zero level (the horizontal black line) and is tabulated in Table 1. (b) The probability distribution functions from p values obtained by applying Welch's t test to the difference between the means $\langle \delta R \rangle_a$ and $\langle \delta R \rangle_b$. The total probability from A_G and N_G is given by the black line. The vertical black line in both panels marks the optimum value of $f_R = 1.120$ with the 2σ uncertainty band ($1.070 \leq f_R \leq 1.155$) shown in grey.

As expected from the correlation coefficients, the residuals are always small for A_G , slightly larger for N_G and sometimes quite considerable for the geomagnetic indices, owing to the limitations in the correlation discussed in section 3.2.

For each of the four calibration sequences and every f_R , the difference $\langle \delta \rangle_a - \langle \delta \rangle_b$ is evaluated and their variations with the f_R value are shown in Figure 7a. The difference decreases linearly with f_R in every case, and the optimum value is where $\langle \delta \rangle_a - \langle \delta \rangle_b$ is zero. This happens at the f_R values given in Table 1 which are in the range 1.079 (a 7.9% correction) to 1.117 (a 11.7% correction).

Comparison of the residuals variations in Figure 6b shows that the agreement is much better for some parameters than others. To derive an overall correction factor that takes this variation into account, we here evaluate the probability p values that $\langle \delta \rangle_a$ and $\langle \delta \rangle_b$ are the same using Welch's t test, as above.

Normalizing over the whole distribution with f_R gives the probability density functions (pdfs) shown in Figure 7b. The pdfs are broader (and thus with a lower peak) for IDV and $IDV(1d)$ because of the lower correlation (the peak r for R and A_G^m being 0.98, whereas the peak r for both $IDV(1d)^{1/n}$ and $IDV^{1/n}$ is 0.84). The broad pdf and the low peak for the geomagnetic data emphasize how much poorer a test of the R calibration is given by the geomagnetic data than is given by the sunspot data. Even for geomagnetic activity indices responding primarily to solar EUV effects on the thermosphere, geomagnetic activity is an indirect indicator of sunspot number, and we concur with *Mursula et al.* [2009] who pointed out that geomagnetic data are not a sufficiently precise measure of sunspot numbers to make them valid for calibration.

We here combine the probabilities by taking their product to generate the pdf in black. The optimum f_R is then taken as the peak of this combined distribution and its uncertainty evaluated at the 2σ level. Table 1 gives

Table 1. Comparison of Results for Three Different Options^a

Data Source	Using A_G , R_G , $IDV(1d)$, and IDV		Using A_G and R_G
	RGO and SOON	RGO Only	RGO and SOON
Optimum f_R	1.116	1.128	1.116
Lower 2σ limit of f_R	1.080	1.091	1.081
Upper 2σ limit of f_R	1.142	1.163	1.148
f_R from A_G	1.116	1.142	1.116
f_R from N_G	1.117	1.126	1.117
f_R from IDV	1.103	1.103	-
f_R from $IDV(1d)$	1.079	1.079	-
$p(f_R = 0.000)$	9.508×10^{-9}	3.964×10^{-9}	3.709×10^{-8}
$p(f_R = 1.200)$	2.711×10^{-6}	1.923×10^{-4}	1.558×10^{-5}

^aThe first column uses all four calibration data series (A_G , R_G , $IDV(1d)$, and IDV); the second uses all four, but only the RGO sunspot group data are used. The third column uses only the sunspot group data. For each, the optimum, maximum, and minimum (at the 2σ level) values of f_R are given along with the values from A_G , R_G , $IDV(1d)$, and IDV in isolation. Also given are the probabilities that no correction is required, $p(f_R = 0.000)$, and that the correction required is 20%, $p(f_R = 1.200)$.

three different combinations of the data. In the first column, all four calibration data series are used. In the third column, we only used the sunspot group data (A_G and N_G) and use the geomagnetic data as an independent check. Because we think it is important to maintain the full independence of the sunspot and geomagnetic data, this is our preferred option. In the second column we search for any influence of the calibration of the SOON and RGO data on the result by only using RGO data. As a check, we have also applied the correlation method (see section 2.1) to define the optimum f_R values. The optimum f_R obtained by correlating with both A_G and N_G was found to be slightly larger, but the uncertainty band around it is broader because fewer data were used.

The combined pdf for the preferred option is shown by the black line in Figure 7b, and the vertical black line gives the optimum f_R ($= 1.116$, i.e., a 11.6% correction), and the shaded grey area gives the 2σ uncertainty band (i.e., 95% of the combined pdf lies in this band) which is between $f_R = 1.081$ (an 8.1% correction) and $f_R = 1.148$ (a 14.8% correction). Although the geomagnetic indices give slightly lower values, including them makes little difference because their probability distributions are so much broader. Using only the RGO data (i.e., we omit the SOON data for 1976–2012) increases the values, but only slightly, so the intercalibration of RGO and SOON is not strongly influencing the results.

Table 1 also quantifies the probabilities that f_R is as small as unity and as large as 1.20. Both are exceedingly small. Therefore, we can discount the possibility that no correction is needed; however, we can equally discount the possibility that it is as large as 20% as was advocated by Svalgaard [2011].

5. Conclusions

We have studied the putative discontinuity in the international sunspot number record in around 1 January 1946. Our results confirm the conclusion of Svalgaard [2011] that the discontinuity is present in the widely used data set that is available from most data centers. However, Svalgaard's estimate that a 20% correction is here shown to be an overestimate. Using the RGO/SOON spot area, Svalgaard [2011] actually found an 18% correction (it is not clear from his paper how this was quantified) that he rounded up to 20%, but both values are far out on the far tail of the probability distribution found here. We suggest preferred value should be based on sunspot data alone, maintaining its independence from geomagnetic data: we have demonstrated problems with early *IDV* data caused by a failure to maintain that independence. Hence, we recommend an optimum correction of 11.6%, with the 2σ uncertainty range being 8–14%. The geomagnetic data provide broad confirmation of this with *IDV* giving and optimum correction of 10.3% and *IDV*(1d) giving 7.9%.

Acknowledgments

We thank a great many scientists who made, recorded, and digitized the solar and geomagnetic observations employed in this paper. The international sunspot number data were obtained from WDC-SILSO, Royal Observatory of Belgium, Brussels (<http://sidc.be/silso/home/>), and the Greenwich/SOON data from D. Hathaway, NASA/Marshall Space Flight Center (<http://solarscience.msfc.nasa.gov/greenwch.shtml>), and the geomagnetic indices are given in the cited papers. We thank the UK Natural Environment Research Council (NERC) for the provision of a Ph.D. studentship for L.B., and subsequent support under NERC grant NE/J024678/1.

Yuming Wang thanks the reviewers for their assistance in evaluating this paper.

References

- Aparicio, A. J. P., J. M. Vaquero, and M. C. Gallego (2012), The proposed "Waldmeier discontinuity": How does it affect to sunspot cycle characteristics?, *J. Space Weather Space Clim.*, 2, UNSP A12, doi:10.1051/swsc/2012012.
- Balmaceda, L., S. K. Solanki, N. Krivova, and S. Foster (2009), A homogeneous database of sunspot areas covering more than 130 years, *J. Geophys. Res.*, 114, A07104, doi:10.1029/2009JA014299.
- Baranyi, T., L. Gyori, A. Ludmány, and H. E. Coffey (2001), Comparison of sunspot area data bases, *Mon. Not. Roy. Astron. Soc.*, 323, 223–230, doi:10.1046/j.1365-8711.2001.04195.x.
- Bartels, J. (1932), Terrestrial-magnetic activity and its relations to solar phenomena, *Terr. Magn. Atmos. Electr.*, 37, 1–52, doi:10.1029/TE037i001p00001.
- Clette, F., D. Berghmans, P. Vanlommel, R. A. M. Van der Linden, A. Koeckelenbergh, and L. Wauters (2007), From the Wolf number to the International Sunspot Index: 25 years of SIDC, *Adv. Space Res.*, 40(7), 919–928, doi:10.1016/j.asr.2006.12.045.
- Cliver, E. W., F. Clette, and L. Svalgaard (2013), Recalibrating the Sunspot Number (SSN): The SSN Workshops, *Cent. Eur. Astrophys. Bull.*, 37, 401–416.
- Erwin, E. H., H. E. Coffey, W. F. Denig, D. M. Willis, R. Henwood, and M. N. Wild (2013), The Greenwich photo-heliographic results (1874–1976): Initial corrections to the printed publications, *Sol. Phys.*, 288(1), 157–170, doi:10.1007/s11207-013-0310-z.
- Foster, S. S. (2004), Reconstruction of solar irradiance variations, for use in studies of global climate change: Application of recent SoHO observations with historic data from the Greenwich observations, PhD dissertation, Univ. of Southampton, U. K.
- Hathaway, D. H. (2010), The solar cycle, *Living Rev. Sol. Phys.*, 7, 1. [Available at <http://www.livingreviews.org/lrsp-2010-1>] (cited on 14 January 2012).
- Hoyt, D. V., and K. H. Schatten (1994), The one hundredth year of Rudolf Wolf's death: Do we have the correct reconstruction of solar activity?, *Geophys. Res. Lett.*, 21(18), 2067–2070, doi:10.1029/94GL01698.
- Hoyt, D. V., and K. H. Schatten (1998), Group sunspot numbers: A new solar activity reconstruction, *Sol. Phys.*, 181(2), 491–512, doi:10.1023/A:1005056326158.
- Lefevre, L., and F. Clette (2014), Survey and merging of sunspot catalogs, *Sol. Phys.*, 289(2), 545–561, doi:10.1007/s11207-012-0184-5.
- Leussu, R., I. G. Usoskin, and K. Mursula (2013), Inconsistency of the Wolf sunspot number series around 1848, *Astron. Astrophys.*, 559, A28, doi:10.1051/0004-6361/201322373.
- Lockwood, M. (2002), An evaluation of the correlation between open solar flux and total solar irradiance, *Astron. Astrophys.*, 382, 678–687, doi:10.1051/0004-6361/20011666.
- Lockwood, M. (2003), Twenty-three cycles of changing open solar flux, *J. Geophys. Res.*, 108(A3), 1128, doi:10.1029/2002/JA009431.

- Lockwood, M. (2013), Reconstruction and prediction of variations in the open solar magnetic flux and interplanetary conditions, *Living Rev. Sol. Phys.*, *10*, 4, doi:10.12942/lrsp-2013-1. [Available at <http://www.livingreviews.org/lrsp-2013-4>.] (Accessed on 14 January 2014).
- Lockwood, M., and M. J. Owens (2014), Centennial variations in sunspot number, open solar flux and streamer belt width: 3. Modelling, *J. Geophys. Res. Space Physics*, doi:10.1002/2014JA019973.
- Lockwood, M., M. J. Owens, L. Barnard, C. J. Davis, and F. Steinhilber (2011), The persistence of solar activity indicators and the descent of the Sun into Maunder minimum conditions, *Geophys. Res. Lett.*, *38*, L22105, doi:10.1029/2011GL049811.
- Lockwood, M., L. Barnard, H. Nevanlinna, M. J. Owens, R. G. Harrison, A. P. Rouillard, and C. J. Davis (2013), Reconstruction of geomagnetic activity and near-Earth interplanetary conditions over the past 167 yr—Part 1: A new geomagnetic data composite, *Ann. Geophys.*, *31*, 1957–1977, doi:10.5194/angeo-31-1957-2013.
- Lockwood, M., H. Nevanlinna, M. Vokhmyanin, D. Ponyavin, S. Sokolov, L. Barnard, M. J. Owens, R. G. Harrison, A. P. Rouillard, and C. J. Scott (2014a), Reconstruction of geomagnetic activity and near-Earth interplanetary conditions over the past 167 years: 3. Improved representation of solar cycle 11, *Ann. Geophys.*, *32*, 367–381, doi:10.5194/angeo-32-367-2014.
- Lockwood, M., H. Nevanlinna, L. Barnard, M. J. Owens, R. G. Harrison, A. P. Rouillard, and C. J. Scott (2014b), Reconstruction of geomagnetic activity and near-Earth interplanetary conditions over the past 167 years: 4. Near-Earth solar wind speed, IMF, and open solar flux, *Ann. Geophys.*, *32*, 383–399, doi:10.5194/angeo-32-383-2014.
- Lockwood, M., M. J. Owens, and L. Barnard (2014c), Centennial variations in sunspot number, open solar flux and streamer belt width: 2. Comparison with geomagnetic data, *J. Geophys. Res. Space Physics*, doi:10.1002/2014JA019972.
- Mursula, K., I. Usoskin, and O. Yakovchouk (2009), Does sunspot number calibration by the “magnetic needle” make sense?, *J. Atmos. Sol. Terr. Phys.*, *71*(17/18), 1717–1726, doi:10.1016/j.jastp.2008.04.017.
- Owens, M. J., and R. J. Forsyth (2013), The heliospheric magnetic field, *Living Rev. Sol. Phys.*, *10*, 5. [Available at <http://www.livingreviews.org/lrsp-2013-5>.] (cited on 10th March 2014).
- Owens, M., and M. Lockwood (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:10.1029/2011JA017193.
- Satterthwaite, F. E. (1946), An approximate distribution of estimates of variance components, *Biometrics Bull.*, *2*, 110–114, doi:10.2307/3002019.
- Solanki, S., M. Schüssler, and M. Fligge (2000), Secular evolution of the Sun’s magnetic field since the Maunder minimum, *Nature*, *400*, 445–446, doi:10.1038/35044027.
- Svalgaard, L. (2011), How well do we know the sunspot number?, *Proc. Int. Astron. Union*, *7*, 27–33, doi:10.1017/S1743921312004590.
- Svalgaard, L., and E. W. Cliver (2010), Heliospheric magnetic field 1835–2009, *J. Geophys. Res.*, *115*, A09111, doi:10.1029/2009JA015069.
- Usoskin, I. G. (2013), A history of solar activity over millennia, *Living Rev. Sol. Phys.*, *5*, 3. [Available at <http://www.livingreviews.org/lrsp-2013-5>.] (cited on 28 February 2014).
- Welch, B. L. (1947), The generalization of “Student’s” problem when several different population variances are involved, *Biometrika*, *34*(1–2), 28–35, doi:10.1093/biomet/34.1-2.28.
- Willis, D. M., H. E. Coffey, R. Henwood, E. H. Erwin, D. V. Hoyt, M. N. Wild, and W. F. Denig (2013), The Greenwich photo-heliographic results (1874–1976): Summary of the observations, applications, datasets, definitions and errors, *Sol. Phys.*, *288*(1), 117–139, doi:10.1007/s11207-013-0311-y.