Centennial variations in sunspot number, open solar flux and streamer belt width: 2. Comparison with the geomagnetic data

Article

Published Version

Creative Commons: Attribution 3.0 (CC-BY)

Open Access


It is advisable to refer to the publisher's version if you intend to cite from the work. See Guidance on citing.
To link to this article DOI: http://dx.doi.org/10.1002/2014JA019972

Publisher: American Geophysical Union

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the End User Agreement.

www.reading.ac.uk/centaur
CentAUR
Central Archive at the University of Reading
Reading's research outputs online
Centennial variations in sunspot number, open solar flux, and streamer belt width: 2. Comparison with the geomagnetic data

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

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

Abstract We investigate the relationship between interdiurnal variation geomagnetic activity indices, IDV and IDV(1d), corrected sunspot number, \( R_C \), and the group sunspot number \( R_G \). \( R_C \) uses corrections for both the “Waldmeier discontinuity,” as derived in Paper 1, and the “Wolf discontinuity” revealed by Leussu et al. (2013). We show that the simple correlation of the geomagnetic indices with \( R_C \) or \( R_G \) masks a considerable solar cycle variation. Using IDV(1d) or IDV to predict or evaluate the sunspot numbers, the errors are almost halved by allowing for the fact that the relationship varies over the solar cycle. The results indicate that differences between \( R_C \) and \( R_G \) have a variety of causes and are highly unlikely to be attributable to errors in either \( R_G \) alone, as has recently been assumed. Because it is not known if \( R_C \) or \( R_G \) is a better predictor of open flux emergence before 1874, a simple sunspot number composite is suggested which, like \( R_G \), enables modeling of the open solar flux for 1610 onward in Paper 3 but maintains the characteristics of \( R_C \).

1. Introduction

In this second paper of a series of three, we study the relationship between geomagnetic interdiurnal range indices and sunspot numbers in detail. We investigate the degree to which sunspot data are consistent with the geomagnetic data and aim to define sunspot data sequences to use as inputs to the open solar flux model prediction in Paper 3 [Lockwood and Owens, 2014].

As discussed in Paper 1 [Lockwood et al., 2014c], there are two main long-term indices in common use to quantify sunspot activity; the sunspot number and the group sunspot number data sequences which extend back to 1749 and 1610, respectively, in monthly means. We here start from the sunspot number \( R \) sequence published by the World Data Centre for the Sunspot Index and Long-term Solar Observations (WDC-SILSO) of the Royal Observatory of Belgium, Brussels, which is a composite of the international, Zürich, and Wolf sunspot numbers. The international number has been compiled from data generated by a number of observers (currently numbering 86 in 29 different countries) from 1 January 1981 onward. This uses the same algorithm as, and is a continuation of, the daily Zürich sunspot numbers, \( R_z \), which were based upon observations made at Zürich and its two branch stations in Arosa and Locarno. Only one observer was used to make each daily \( R_z \) estimate (who was the highest ranked observer available on that day in a hierarchy ordered by perceived reliability). The “Waldmeier discontinuity” discussed in Paper 1 relates to a putative change in observing and processing practices for compiling \( R_z \) around 1945.

As discussed by Clette et al. [2007], the basic sunspot number formulation was first used in 1849 by Rudolf Wolf, Director of the Zürich Observatory. He was succeeded by his former assistant Alfred Wolfer, who introduced an important change in 1882. In generating his sunspot number, \( R_W \), Wolf did not count the smallest spots when quantifying the number of individual spots, \( N_S \), in order to try to maintain compatibility with earlier observers. However, this requires a subjective choice of which spots to include. Wolfer abandoned this practice and his new procedure was calibrated against Wolf’s sunspot numbers over a 16 year interval (1877–1892). This yielded a seemingly constant scaling factor of 0.6 that is still used today to scale the sunspot observations to the pre-1882 Wolf sunspot number, \( R_W \). In addition, Wolf had become so convinced of a linear correlation between his geomagnetic and sunspot data that he recalibrated his sunspot numbers for before 1849 upward by about 20%. An additional problem is that the data before 1849 are sparse, leading to greater uncertainties [Usoskin, 2013]. Another inhomogeneity in the data series arises in 1818, before when only monthly values have been compiled, whereas after then the basic data available are daily. Annual means of \( R_W \) are available from 1700 [Chernosky and Hagan, 1958], but they are not generally regarded as reliable because the data are so sparse and highly interpolated and consequently these are rarely used [Usoskin, 2013].
Leussu et al. [2013] have recently used a homogeneous series of recently digitized sunspot drawings by Schwabe [Arlt et al., 2013] to study the calibration of the Wolf sunspot numbers, \( R_W \). They find a systematic error in the scaling of \( R_W \) before 1849, calling for a 20% reduction in values of \( R \) for all years before 1849. This almost certainly is reversing Wolf’s 20% upward recalibration of sunspot numbers to agree with geomagnetic activity data, a practice that is now generally thought to be both unreliable and undesirable [Mursula et al., 2009, Lockwood et al., 2014c]. Hence, we will refer to the correction called for by the work of Leussu et al. as the “Wolf discontinuity.”

The group sunspot number series was introduced by Hoyt and Schatten [1994, 1998] and is extremely valuable because it extends back to before the Maunder minimum in sunspot activity. However, it has to be compiled from observations with sparse availability for the early years (in equation (2) of Paper 1 is small and some observers may not have been active throughout any 1 year). Usoskin et al. [2003a] studied the effect of this and made some minor corrections, and other corrections have been made, based on additional historic observations that have come to light [Vaquero et al., 2011; Vaquero and Trigo, 2014]. The Royal Greenwich Observatory (RGO)/Solar Observing Optical Network (SOON) sunspot group data discussed in Paper 1 cover 1874 to the present and form the backbone of the \( R_C \) data sequence, and the correlation demonstrated in Paper 1 is used to ensure that it is very similar to the international sunspot number since 1874. However, it is well known that international/Zürich/Wolf composite and group sunspot numbers diverge as one goes back in time before this date, as indeed was noted by Hoyt and Schatten [1994, 1998] when they first derived the \( R_Q \) sequence (e.g., Hathaway et al., 2002).

Paper 1 quantified the magnitude of the Waldmeier discontinuity to be 11.6%. In addition, Leussu et al. [2013] have quantified the Wolf discontinuity to be 20%. We here implement both these corrections before establishing the relationship between the geomagnetic variation and sunspot number and how it varies over the solar cycle. To ensure that there is no confusion, we here call the corrected international/Zürich/Wolf sunspot number composite \( R_C \) where \( R_C = R \) for 1946 and after, \( R_C = 1.116 R \) for 1849–1945, and \( R_C = (1.116 \times 0.8)R \) for 1848 and before, where \( R \) is the sunspot number composite, as published by WDC-SILSO. In the present paper we also suggest a simple extension of the resulting \( R_C \) back to 1610 using 1.3\( R_C \). (Note that this is not, in any sense, a recalibration of \( R_C \); rather, it is a means of extending the \( R_C \) data series back to the Maunder minimum that is greatly preferable to using the sparse, highly interpolated early annual means of the Wolf number).

Svalgaard and Cliver [2005] introduced the IDV index based on the interdiurnal variability \( u \) index of Bartels [1932]. The major difference between \( u \) and IDV is that the latter is based on near-midnight values only, whereas the \( u \) index used whole-day means. Svalgaard and Cliver [2007] did not see this as a significant difference because they used Bartels’ \( u \) index data to extend the IDV sequence back to 1835, which is just 3 years after Gauss’ establishment of the first magnetometer station in Göttingen. The data used by Svalgaard and Cliver [2007] to extend IDV to before 1880 were compiled by Bartels, but it is not a homogeneously constructed index, being compiled in a different way after 1872 to before then. Bartels notes that before 1872, no proper data to generate an interdiurnal index were available to him and so he used other correlated measures of the diurnal variation as proxies. Bartels himself stresses that his \( u \) values before 1872 are “more for illustration than for actual use” and describes data for 1835–1847 as “least reliable,” 1847–1872 as “better,” and 1872–1930 as “satisfactory.” Given that Bartels does not include his data before 1872 in his satisfactory classification, it is not just a semantic point that he regarded the data before 1872 as “unsatisfactory.” Svalgaard and Cliver [2007] carried out some tests to justify employing the \( u \) proxy data for before 1872 which enabled them to make a reconstruction of geomagnetic activity back to 1835. However, Figure 5 of Paper 1 [Lockwood et al., 2014c] shows that the early IDV data have different characteristics to the later data. Specifically, the early IDV index was much more similar to the sunspot number data than at later times strongly suggesting that the independence of the two data sets had not been maintained and that at some point calibration of one using the other had taken place. This may partially have come about through Wolf’s use of geomagnetic data to recalibrate \( R_W \), but as shown in Paper 1, other geomagnetic data do not show the same error as IDV. This is a pitfall we remain conscious to avoid in this paper.

Lockwood et al. [2013a] introduced the IDV(ID) index because of the following: (1) they were concerned about the overstrong correlation between \( u \) (and hence early IDV) and sunspot numbers in the early years and (2) because they found that the dependence on interdiurnal variation data on interplanetary parameters depended on station latitude (at all latitudes, not just near the auroral oval) and because IDV was compiled from an evolving mix of stations, its dependence on interplanetary parameters in the past would be different from that in the space age. These authors also returned to Bartel’s original concept of using whole-day means which has advantages
in suppressing both instrumental and geophysical noise: indeed, the range of correlations for stations at different latitudes is found to be as great as that for one station at different UTs, and hence, 24 stations of IDV data are required to achieve the same noise suppression as one station giving IDV(1d). This has the advantage of allowing IDV(1d) to be compiled from just one magnetic observatory in the same geographical region, avoiding artifacts caused by the response to interplanetary variations varying with the location of the observatory. Lockwood et al. [2013a] were able to extend the IDV(1d) data sequence back to 1845 using data from the Helsinki observatory. Svalgaard [2014] pointed out a calibration error in the Helsinki data that applied to 7 years during cycle 11, and Lockwood et al. [2014a] have studied the optimum correction method and implemented it to allow for this. It is important to note that although Svalgaard [2014] used RG to identify the calibration change in the Helsinki data, the correction implemented by Lockwood et al. [2014a] did not employ any form of sunspot number and so the full independence of the geomagnetic and sunspot number data sequences was maintained. From modeling using R to quantify emergence rate, Wang et al. [2005] found an approximate relationship between open solar flux and R, with an exponent of n near 0.5. However, these authors also noted that the correlation was high for photospheric flux but poor for open solar flux. Svalgaard and Cliver [2005] found that their IDV index correlated with R. They also noted that IDV correlates with the near-Earth interplanetary magnetic field (IMF) B with very little influence of the solar wind speed (see Lockwood [2013]) so Svalgaard and Cliver [2005] concluded that B also varies with R. As noted by Lockwood et al. [2014a], the peak correlation coefficient between the sunspot number to the power n, Rn, and the observed IMF is r = 0.84 for n = 0.4, with significance (computed against the AR-1 autoregressive “red-noise” model) of S = 96.3%. Hence, the simultaneous Rn “explains” about 70% of the overall observed IMF variation. Paper 1 shows the peak correlations for IDV and IDV(1d) are 0.85 and 0.83 for n = 0.54 and n = 0.69, respectively. Both are more statistically significant than the correlation with the IMF data (S > 99.9%) because of the larger number of data samples. We here investigate these correlations of IDV and IDV(1d) with sunspot numbers in more detail.

2. The Nature of the Relationship Between Interdiurnal Variation Indices and Sunspot Number

Figure 1a shows a scatterplot of annual means for 1874 to 2012 (inclusive) of the corrected international sunspot number, Rc, as a function of IDV(1d)^1/n, where n is the best fit exponent (n = 0.69), derived in Paper 1 and Rc employs the optimum corrections fR1 and fR2 discussed in the text. (a) Takes no account of the solar cycle phase, whereas in (b) the data are sorted into seven equal-sized bins of solar cycle phase, Φ, derived from the RG/SOON data using the mean heliographic latitude of sunspot groups and the method of Owens et al. [2011]. Points for each bin are color coded and fitted with a linear regression constrained to pass through the point Rc = 0, IDV(1d) = 2.6 (IDV(1d)^1/n = 4); the linear regression fits have slope F and are shown by the lines using the same color coding.

Figure 1. Scatterplots of annual means for 1874 to 2012 (inclusive) of the corrected international sunspot number, Rc, as a function of IDV(1d)^1/n, where n is the best fit exponent (n = 0.69), derived in Paper 1 and Rc employs the optimum corrections fR1 and fR2 discussed in the text. (a) Takes no account of the solar cycle phase, whereas in (b) the data are sorted into seven equal-sized bins of solar cycle phase, Φ, derived from the RG/SOON data using the mean heliographic latitude of sunspot groups and the method of Owens et al. [2011]. Points for each bin are color coded and fitted with a linear regression constrained to pass through the point Rc = 0, IDV(1d) = 2.6 (IDV(1d)^1/n = 4); the linear regression fits have slope F and are shown by the lines using the same color coding.
Figure 2. Variation of the slope of the regressions lines in Figure 1b, F(Φ) as a function of Φ. The points are the best fit values for the seven bins (shown using the same color coding as Figure 1b) and the line is an interpolation using splines fitting. Points are shown for the mean Φ of the data in each bin. The best linear regression is fit to the annual mean with the IDV for the best fit exponent n = 0.69 derived from Figure 5 of Paper 1. The data are for 1874–2012. Note that the factor f_{R1} is based on the RGO/SOON sunspot data only and the factor f_{R2} is based on reanalysis of Schwabe’s sunspot drawings, and so neither has been influenced by any geomagnetic data. The best fit linear regression line is shown, fitted using a Bayesian least squares fit. At first sight there appears to be a simple relationship. However, Figure 1b shows that this correlation is not all it seems and that it masks a strong solar cycle variation. The data are here divided into seven equal size bins according to the solar cycle phase, Φ. This phase is determined by the method of Owens et al. [2011] which exploits the regular solar cycle (butterfly wing) variation of the mean latitude of sunspot groups from the RGO/SOON data: Φ is defined as zero (2π) at the start of each new cycle when the mean latitude of sunspots increases suddenly. Points are color-coded by their Φ bin and for each a linear regression is fitted, each constrained to pass through a set common origin (and shown in Figure 1b using the same color scheme). The fits all pass through the point IDV(1d) = 2.6 (which gives IDV(1d)^{1/n} = 4), R_C = 0. This point was derived by varying the intercept and evaluating the probability density functions (pdfs) for each fit: these were very similar in form and taking the peak of the product of the pdfs gives the optimum intercept value for all seven fits. It can be seen that the relationship between IDV(1d)^{1/n} and R_C depends strongly on Φ with a coherent solar cycle variation giving a much larger slope, f, as sunspot maximum (the light blue and green lines) than at sunspot minimum (the red/brown and blue lines). Hence, using an overall fit shown in Figure 1a would introduce systematic differences at both sunspot minimum and sunspot maximum.

The best fit slope F(Φ) = dR_C/dIDV(1d)^{1/n}, of each linear regression fit is shown in Figure 2, using the same color coding as Figure 1b, as a function of the sunspot cycle phase Φ. The black line is an interpolation using a Piecewise Cubic Hermite Interpolating Polynomial fit that is here used to compute F for any value of Φ. Figure 3a shows the scatterplot of the correction factor f_{R2} = 0.8 for the Wolf discontinuity, as derived by Leussu et al. [2013], against the simultaneous annual means of IDV(1d)^{1/n}, for the best fit exponent n = 0.69 derived from Figure 5 of Paper 1. The data are for 1874–2012. Figure 3b shows the corrected sunspot number R_C using the same color coding as Figure 1b, as a function of the sunspot cycle phase Φ. The point was derived by varying the intercept and evaluating the probability density functions (pdfs) from the data, divided into seven equal size bins according to the solar cycle phase, Φ. This phase is determined by the method of Owens et al. [2011] which exploits the regular solar cycle (butterfly wing) variation of the mean latitude of sunspot groups from the RGO/SOON data: Φ is defined as zero (2π) at the start of each new cycle when the mean latitude of sunspots increases suddenly. Points are color-coded by their Φ bin and for each a linear regression is fitted, each constrained to pass through a set common origin (and shown in Figure 1b using the same color scheme). The fits all pass through the point IDV(1d) = 2.6 (which gives IDV(1d)^{1/n} = 4), R_C = 0. This point was derived by varying the intercept and evaluating the probability density functions (pdfs) for each fit: these were very similar in form and taking the peak of the product of the pdfs gives the optimum intercept value for all seven fits. It can be seen that the relationship between IDV(1d)^{1/n} and R_C depends strongly on Φ with a coherent solar cycle variation giving a much larger slope, f, as sunspot maximum (the light blue and green lines) than at sunspot minimum (the red/brown and blue lines). Hence, using an overall fit shown in Figure 1a would introduce systematic differences at both sunspot minimum and sunspot maximum.

Figure 3. (a) Scatterplot of R_C/F(Φ), determined using the R_C and Φ value for each annual mean with the F(Φ) value interpolated using Figure 2, as a function of IDV(1d)^{1/n} for the optimum n = 0.69. The black line is the best linear regression fit R_C/F(Φ) = s IDV(1d)^{1/n} + c for the best fit coefficients of s = 1.000 and c = 3.966. (b) The corrected sunspot number R_C as a function of its estimate from observed IDV(1d), R_{IDV(1d)} = F(Φ) × [s IDV(1d)^{1/n} + c]. In both panels the annual data points are color coded according to their solar cycle phase, Φ.
Figure 4 shows a scatterplot of the annual values of $R_C/F(\Phi)$ against $IDV(1d)^{1/n}$ obtained this way. It can be seen that the scatter has been reduced by the use of $F(\Phi)$ compared to Figure 1a and that the fit now applies at all phases of the solar cycle with no systematic biases of the residual with solar cycle phase. The black line is the best linear regression of slope $s$ and intercept $c$, using which the value of $IDV(1d)$ can be used to compute the best fit to $R_C$, wherever $R_{IDV}(1d) = F(\Phi) \times [s^{IDV(1d)} + c]$, where the best fit coefficients are $s = 1.000$ and $c = -3.966\ nT^{1/n}$. Note that $s$ is unity, as it should be from the definition of $F(\Phi)$. A scatterplot of $R_{IDV(1d)}$ against $R_C$ is shown in Figure 3b. By allowing for the phase of the solar cycle, $\Phi$, the correlation coefficient has been raised from $r = 0.87$ ($r^2 = 0.75$) in Figure 1a to $r = 0.95$ ($r^2 = 0.90$) in Figure 3b and so this enables us to “explain” a further 15% of the variation in $R_C$. The RMS deviation between observed and fitted $R_C$ values has been reduced from 22.4 to 13.8.

This is not just a property of the $IDV(1d)$ index: the $IDV$ index behaves in exactly the same way. Without allowance for $\Phi$, the correlation of $R_C$ with against $IDV^{1/n}$ peaks at $r = 0.87$ for $n = 0.54$ ($r^2 = 0.75$), whereas $R_{IDV}$ and $R_C$ give $r = 0.94$ ($r^2 = 0.88$) and the RMS deviation between observed and fitted $R_C$ values is reduced from 22.7 to 14.5 by including an allowance for $\Phi$.

Hence, if using $IDV(1d)$ or $IDV$ to predict or evaluate the sunspot number $R_C$, the prediction errors are almost halved by allowing for the fact the relationship varies over the solar cycle. Because $R_C$ and $R_G$ are very similar over much of the interval of the geomagnetic observations, very similar results were obtained by fitting to $R_G$ rather than $R_C$.

3. Comparison of Sunspot Number Estimates

Figure 4 compares the time series of the corrected sunspot numbers ($R_C$, in black in top) and the group sunspot numbers ($R_G$, in black in bottom) with values estimated with allowance for $\Phi$ from $IDV(1d)$ and $IDV, R_{IDV(1d)}$ and $R_{IDV}$ which are shown in both panels by the mauve and the blue lines, respectively. Agreement is generally good using both geomagnetic indices—the slightly higher correlation (and lower RMS fit residual) obtained using $IDV(1d)$ is mainly because $IDV$ tends to give some solar cycle maxima that are too large, compared to the maxima in both $R_C$ and $R_G$. The white and gray stripes show even- and odd-numbered solar cycles, respectively; the cycle numbers being given along the top of each panel.
Paper 1 discusses the sunspot number variation after the start of the RGO data in 1874 (solar cycles 12–13), and we here concentrate in the behavior before then, namely, cycles 9–11. Figure 4 shows that the sunspot number variation predicted by IDV(1d) is consistent with the RC over solar cycle 11, whereas the group number is considerably smaller. The value for IDV is larger than all other values for this cycle. R IDV(1d) and R IDV both increase with time over cycles 9–11 such that although both are much closer to RC for cycle 11, for cycle 9 they both lie midway between RG and RC. In this comparison, it should be remembered that the higher values of R IDV(1d) in cycle 11 arise from recalibration of the Helsinki data proposed by Svalgaard [2014]. Svalgaard’s correction was based on sunspot numbers but the IDV index implemented here is that of Lockwood et al. [2014a] who used only geomagnetic data (from Helsinki and St. Petersburg) to make this correction, thereby maintaining the independence of the sunspot and geomagnetic data.

Table 1 gives the means over the sunspot cycles of both RC and RG and their ratio: it shows that cycle 11 gives <RC>/<RG> of 1.45 which is a maximum for after the Dalton minimum but this ratio declines as we go back in time and is essentially unity for cycle 7 (immediately after the Dalton minimum). Going to yet earlier cycles, <RC>/<RG> rises again, reaching a peak of 1.75 in cycle 5 before falling again to 1.30 for cycle 1.

4. Discussion

Figure 5a shows the annual means of RG than RC (in red and blue), and Figure 5b gives the 22 year running means <RG>22 and <RC>22 (in the same colors) and Table 1 gives the means over each solar cycle (minimum to minimum) and their ratio. Cliver et al. [2013] study the ratio of annual values RG/RC values (where R is the WDC-SILSO sunspot number composite), omitting years with R below 5 to avoid generating large values. They define a discontinuity around 1882 which they attribute to an error in the calibration of RG. They then assume that the R calibration is constant (so all differences between R and RG are assumed to be erroneous and all errors are assumed to be in RG), thereby generating a new sunspot number sequence. In addition, they use an overly large estimate for fa1 of 1.2 (remember that Paper 1 finds an optimum value of 1.116 and that the probability of fa1 = 1.2 is miniscule), and they neglect the Wolf discontinuity (i.e., they use fR2 of 1). Allied together these three decisions have the effect of making pre-1885 values of sunspot number much larger and so make the long-term drift in sunspot numbers appear much smaller [see Svalgaard, 2011].

Table 1. Analysis of Minimum-to-Minimum Sunspot Cycle Means for Solar Cycles 1–16

<table>
<thead>
<tr>
<th>Cycle Number</th>
<th>Start Year</th>
<th>End Year</th>
<th>&lt;RC&gt;</th>
<th>&lt;RG&gt;</th>
<th>&lt;RC&gt;/&lt;RG&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1754</td>
<td>1765</td>
<td>37.11</td>
<td>28.57</td>
<td>1.30</td>
</tr>
<tr>
<td>2</td>
<td>1765</td>
<td>1775</td>
<td>50.03</td>
<td>43.22</td>
<td>1.16</td>
</tr>
<tr>
<td>3</td>
<td>1775</td>
<td>1783</td>
<td>65.95</td>
<td>39.71</td>
<td>1.66</td>
</tr>
<tr>
<td>4</td>
<td>1783</td>
<td>1798</td>
<td>61.73</td>
<td>34.63</td>
<td>1.49</td>
</tr>
<tr>
<td>5</td>
<td>1798</td>
<td>1810</td>
<td>21.27</td>
<td>12.14</td>
<td>1.75</td>
</tr>
<tr>
<td>6</td>
<td>1810</td>
<td>1823</td>
<td>16.14</td>
<td>10.82</td>
<td>1.49</td>
</tr>
<tr>
<td>7</td>
<td>1823</td>
<td>1833</td>
<td>34.84</td>
<td>33.52</td>
<td>1.04</td>
</tr>
<tr>
<td>8</td>
<td>1833</td>
<td>1843</td>
<td>58.28</td>
<td>50.73</td>
<td>1.15</td>
</tr>
<tr>
<td>9</td>
<td>1843</td>
<td>1856</td>
<td>56.07</td>
<td>41.65</td>
<td>1.35</td>
</tr>
<tr>
<td>10</td>
<td>1856</td>
<td>1867</td>
<td>55.34</td>
<td>41.72</td>
<td>1.33</td>
</tr>
<tr>
<td>11</td>
<td>1867</td>
<td>1878</td>
<td>63.12</td>
<td>43.65</td>
<td>1.45</td>
</tr>
<tr>
<td>12</td>
<td>1878</td>
<td>1889</td>
<td>38.55</td>
<td>30.12</td>
<td>1.28</td>
</tr>
<tr>
<td>13</td>
<td>1889</td>
<td>1901</td>
<td>43.27</td>
<td>40.30</td>
<td>1.07</td>
</tr>
<tr>
<td>14</td>
<td>1901</td>
<td>1913</td>
<td>34.76</td>
<td>32.28</td>
<td>1.08</td>
</tr>
<tr>
<td>15</td>
<td>1913</td>
<td>1923</td>
<td>49.28</td>
<td>49.07</td>
<td>1.00</td>
</tr>
<tr>
<td>16</td>
<td>1923</td>
<td>1933</td>
<td>45.81</td>
<td>46.46</td>
<td>0.99</td>
</tr>
<tr>
<td>17</td>
<td>1933</td>
<td>1944</td>
<td>61.38</td>
<td>59.01</td>
<td>1.04</td>
</tr>
<tr>
<td>18</td>
<td>1944</td>
<td>1954</td>
<td>76.16</td>
<td>72.26</td>
<td>1.05</td>
</tr>
<tr>
<td>19</td>
<td>1954</td>
<td>1964</td>
<td>94.89</td>
<td>86.86</td>
<td>1.09</td>
</tr>
<tr>
<td>20</td>
<td>1964</td>
<td>1976</td>
<td>58.80</td>
<td>58.39</td>
<td>1.01</td>
</tr>
<tr>
<td>21</td>
<td>1976</td>
<td>1986</td>
<td>82.98</td>
<td>83.32</td>
<td>1.00</td>
</tr>
</tbody>
</table>

*In addition to the cycle start and end dates, the cycle means of the corrected international sunspot number, <RC>, and the sunspot group number, <RG>, are given along with the ratio of the two <RC>/<RG>.*
However, Figures 4 and 5 and Table 1 cast major doubt that \( R_C \) values before 1882 are essentially correct and the variations in the ratio \( <R_C>/<R_G> \) are due to errors in \( <R_G> \) alone. The vertical dash-dotted lines in Figure 5a show the times of significant changes in the compilation of the \( R \) (and hence \( R_C \)) composite. The last of these is the change from Zürich to international sunspot numbers. The annual means of \( R_C \) and \( R_G \) in Figure 5a show no relative change across this date and although \( <R_G>_{22} \) values in Figure 5b only extend for a few years after this date, there is no sign of any discontinuity. Before this is the Waldmeier discontinuity in 1945 which sits close to the minimum between cycles 17 and 18, and Table 1 shows that \( <R_C>/<R_G> \) is essentially the same for these two cycles. Hence, any variation between cycles 15 and 21 is not attributable to the Waldmeier discontinuity (which has now been corrected for) and is associated with measurement uncertainties in \( R_C \) or \( R_G \), or both, or with a real variation in the ratio of the numbers of spots to groups (\( N_S/N_G \)) which would influence \( R_C \) and \( R_G \) differently.

The dash-dotted line at 1882 (during cycle 12) marks the change from Wolf to Wolfer’s sunspot counting procedure. Table 1 shows that \( <R_C>/<R_G> \) increases strongly as we go back in time through this date, which can also be seen in Figures 5a and 5b. Hoyt and Schatten [1994, 1998] were aware of this when they compiled \( R_G \) and, quite reasonably, attributed it to inadequate calibration of the known change from Wolf to Wolfer’s method and hence an error in \( R \). On the other hand, Cliver et al. [2013] accept the validity of the 0.6 Wolfer calibration factor as a constant and claim the error is in \( R_G \). The geomagnetic data do give some support to Cliver et al.’s view in that in Figure 4 the sunspot reconstructions for cycle 11 from geomagnetic activity, \( R_{IDV(1d)} \) and \( R_{IDV} \) both agree somewhat better with \( R_C \) than \( R_G \). However, Wolfer’s derivation of the 0.6 factor was based on data from 1877 to 1892, and Table 1 shows that as we go back in time to before 1877, into cycles 10 and then 9, the ratio \( <R_C>/<R_G> \) falls, and Figure 4 shows that the geomagnetic reconstructions lie
midway between $R_C$ and $R_G$ in cycle 9 and indeed the more reliable test is $IDV(1d)$ (because, as shown in Paper 1, it is uncontaminated by sunspot number) and this gives $IDV(1d)$ and that is more similar to $R_C$ than $R_G$ in cycle 9.

Cycles 7 and 8 give $<R_C>/<R_G>$ close to unity because of the correction by Leussu et al. [2013] for the Wolf discontinuity (the dash-dotted line at 1848 during cycle 9). Hence, the relative behavior of $R_C$ and $R_G$ and the work of Leussu et al. [2013] shows that using Wolfer’s constant factor of 0.6 before 1877, as assumed by Cliver et al. [2013] and Svalgaard [2011], is too simplistic and invalid.

Table 1 and Figure 5b show that the corrections for the Wolf and Waldmeirer discontinuities make $R_C$ and $R_G$ more similar in their long-term variation than $R$ and $R_C$. Nevertheless, there are still differences which are considerably greater before 1882 than after that date, and those differences do fluctuate with time. Given the subjectivity of the decisions on which individual spots to include in Wolf’s sunspot numbers before 1882, we agree with the many authors who regard $R_G$ as more robust and homogeneous than $R$ (and hence $R_C$) (see reviews by Usoskin [2013], Hathaway [2010], Hathaway and Wilson [2004], and references therein). However, it is possible that calibration errors in both sequences contribute to these fluctuations at different times and to varying degrees, and we can never be sure that there has not been a long-term variation in sunspot behavior which means that at least a component of the divergences or $R_C$ and $R_G$ are real phenomena rather than measurement artifacts.

In the context of this last point, we here note that recent cycles do not lead us to expect the ratio $R_C/R_G$ to be exactly constant, as shown by Figure 5 and Table 1 (cycles 16 to 21 show a 10% peak-to-peak fluctuation in $<R_C>/<R_G>$. Penn and Livingston [2006] have created much interest in the variability of magnetic fields within sunspots. They deduced a decline over the interval 1998–2005, whereas other authors report only a solar cycle variation with little change from one cycle to the next [Pevtsov et al., 2011]. Nagovitsyn et al. [2012] note that these data can be reconciled by the observation that there has been a change over 1998–2005 in the fraction of small and large sunspots which has the potential to cause different changes in $R$ and $R_G$ and hence to vary the ratio $R/R_G$. In addition, the results of Kilcik et al. [2011, Figure 3] show changes in the ratios $R/N_{GS}$ and $R/N_{GL}$ where $N_{GS}$ is the number of small sunspot groups and $R/N_{GL}$ is the number of large sunspot groups. Specifically, solar cycle means of $R/(N_{GS} + N_{GL})$ are approximately 1.1, 1.2, 1.0, and 0.8 for solar cycles 20–23 so there is a ±20% variation about the mean. Furthermore, there have been changes in the distribution of group sizes with $(N_{GS}/N_{GL}) \approx 2.1, 1.6, 2.0,$ and 1.4 for those cycles which, for a given definition of what constitutes a group, will have different effects on $R$ and $R_G$. Hence, some variations in $R/R_G$ (and hence $R_C/R_G$) are real and should not be attributed to calibration errors.

Separating any calibration drifts in both $R_C$ and $R_G$ and so identifying any effects of real changes in the behavior of the individual spots will require a great deal of careful checking of individual historic sunspot observations that is beyond the scope of this paper. The geomagnetic test reported on here may well be improved and/or extended back toward the ultimate start date of 1832 (the date of the establishment of the first magnetometer station by Gauss) as more digital data become available, but the calibration of that early data must be kept fully independent of sunspot records if it to be of any value. Other tests may be possible, for example, using auroral sighting catalogues.

However, there is one comparison that we can make at this stage, and that is with the heliospheric modulation potential $\phi$ derived from cosmogenic isotopes. Figure 5c shows two examples of $\phi$ reconstructions (22 year means) from $^{10}\text{Be}$ (in green) and $^{14}\text{C}$ (in orange) from the data sets by, respectively, Usoskin et al. [2003b] and Muscheler et al. [2007], as reviewed by Usoskin [2013]. It can be seen that they are not identical, and it must be noted that there are other reconstructions which are different again. Both show considerable similarities to both the smoothed sunspot number records shown in Figure 5b. We here take the arithmetic mean of these two ($\phi_{av} = 0.5(\phi_{10}\text{Be} + \phi_{14}\text{C}$), shown in cyan) and Figure 5b shows (again in cyan) a linear regression of $\phi_{av}$ against smoothed sunspot number for 1880–2000. Several features of this variation from cosmogenic isotopes are also seen in both the $<R_C>$ and $<R_G>$ variations before 1873. Interestingly, the cosmogenic isotope data, like the geomagnetic data, suggest that $R_G$ is more accurate than $R_C$ during solar cycle 11, whereas the opposite appears to be true in cycle 9.

5. Definition of a Usable and Consistent Composite for Modeling Purposes

This paper has discussed the considerable similarities and relatively small differences between various proxies of the solar magnetic activity level. It has been shown that implementing corrections to the sunspot...
number record for the Wolf and Waldmeier discontinuities improves agreement, but we caution that, because they are different indices, one should not expect perfect agreement between sunspot numbers and group numbers. However, thus far, we have not achieved the stated aim of defining a sunspot number (as opposed to group number) variation that can be used as an input into the modeling in Paper 3, which requires extending it back to into or before the Maunder minimum because of the effect of the prior history of sunspot numbers and associated open flux emergence. From the above, we cannot conclude if $R_C$ or $R_G$ before 1874 contains calibration errors, or both or even neither (in which case all the differences would be associated with real changes in sunspot characteristics). Most importantly, for the purposes of Paper 3, we do not know if $R_C$ or $R_G$ better quantifies open flux emergence before 1874. This being the case, the best option is to investigate the implications of both. This is straightforward in the case of $R_G$ because it extends back to before the Maunder minimum, but for $R_C$ we require an extension to earlier years.

We here suggest the simplest algorithm possible. To extrapolate $R_C$ to dates before 1749, we suggest using $1.3R_G$ which applies to solar cycle 1, thereby eliminating any discontinuity at 1749. This extension is shown in Figures 5a and 5b in black. This simple combination is reasonably similar to the average cosmogenic isotope variation. We note there are several refinements to this simple algorithm that could be made (for example, using uncorrected $R_G$ in solar cycle 9 where it appears to agree better with both geomagnetic data and cosmogenic isotopes); however, given the large uncertainties in the calibration of both $R_C$ and $R_G$ that remain, we do not believe that they can be justified. We do not make any claims that this variation in sunspot number is in any way definitive; indeed, we expect more precise variations to emerge as a result of the several ongoing studies of historic sunspot observations to recalibrate both the international and group sunspot number and from improved ice cores yielding better cosmogenic isotope records. However, we do note that this variation is reasonably consistent with both the geomagnetic data and cosmogenic isotope abundances which provide the best independent tests available to us. It is important to stress that we do not mean to suggest, in any way, that this extended $R_C$ is more accurate than $R_G$: indeed, we subscribe to the view that the relative simplicity of $R_G$ means that it is much more robust and homogeneous, even after the corrections used in $R_C$ have been implemented. However, because it remains possible that the differences between $R_C$ and $R_G$ may be real and that $R_C$ may be a better quantifier of open flux emergence, we need to be able to use both as an input to open flux continuity models. Hence, the last paper of this series of three [Lockwood and Owens, 2014] uses both the $R_C$ and extended $R_C$ sequences of sunspot number estimates as an input into a model of open solar flux variation. The out from this model is compared with the reconstructed open solar flux variation which has been generated from the geomagnetic data [Lockwood, 2013; Lockwood et al., 2013a, 2013b, 2014a, 2014b]. The $R_G$ and $R_C$ data sequences are available in the supporting information. The finalized IDV(1d) index data are given in the paper by Lockwood et al. [2014c] and in the supporting information by Lockwood and Owens [2014].

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 (URL: http://sidc.be/silso/home/), the group sunspot data from NGDC, Boulder (URL: http://www.ngdc.noaa.gov/stp/solar/), and the geomagnetic indices and cosmogenic isotope data are given in the cited papers. We thank the UK Natural Environment Research Council (NERC) for the provision of a PhD studentship for LB and subsequent support under NERC grant NE/J024678/1.

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

References


