SH13A-1109

Origins of the Wolf Sunspot Number Series: Geomagnetic Underpinning

Cliver, E W (edward.cliver@hanscom.af.mil), Space Vehicles Directorate, Air Force Research Laboratory.
* Svalgaard, L (leif@leif.org). Easy Tool Kit, Inc.

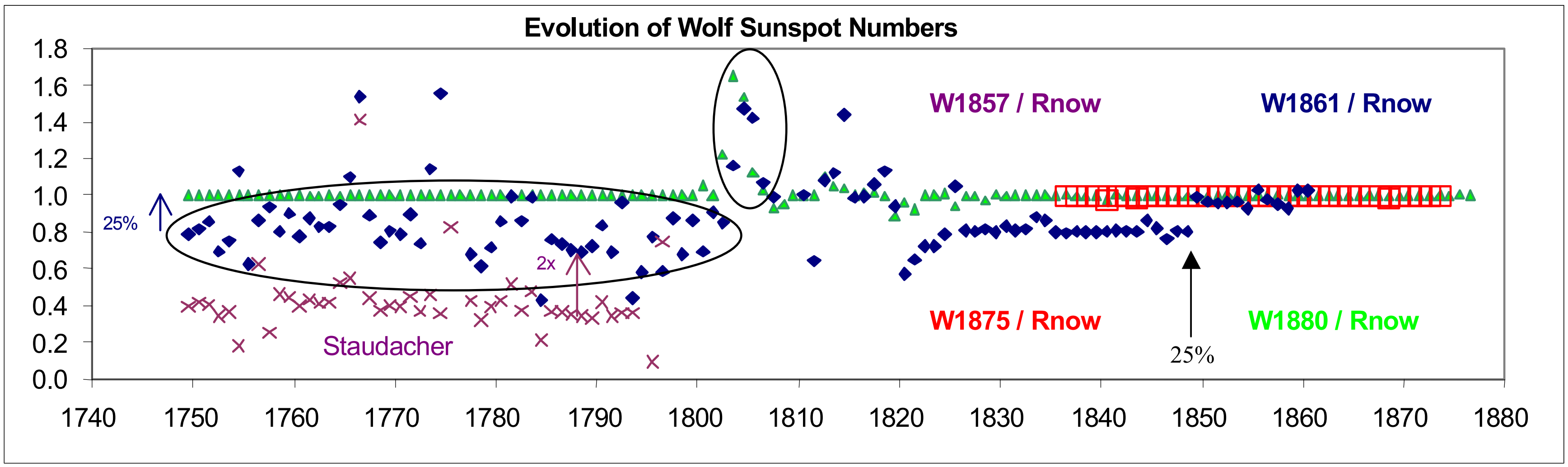
Abstract: The Wolf or International sunspot number (SSN) series is based on the work of Swiss astronomer Rudolf Wolf (1816-1893). Following the discovery of the sunspot cycle by Schwabe in 1843, Wolf culled sunspot counts from journals and observatory reports and combined them with his own observations to produce a SSN series that extended from 1700-1893. Thereafter the SSN record has been maintained by the Zurich Observatory and, since 1981, by the Royal Observatory of Belgium. The 1700-1893 SSN record constructed by Wolf has not been modified since his death. Here we show that Wolf's SSNs were not based solely on reports of sunspots but were calibrated by reference to geomagnetic range observations, which closely track the sunspot number. Nor were these corrections small; for example Wolf multiplied the long series (1749-1796) of sunspot counts obtained by Staudacher by factors of 2.0 and 1.25, in turn, to obtain the numbers in use today. It is not surprising then that a competing SSN series obtained by Hoyt and Schatten based on group sunspot numbers is different, generally lower than that of Wolf. Comparison of the International number with current magnetic range observations indicates that, as Wolf found, the magnetic range (specifically, the average annual range of the Y-component of mid-latitude stations) can be used as an independent check on the validity and stability of the SSN series. Moreover, the geomagnetic range series, which in itself is a long-term proxy of solar FUV emission, can be used to resolve discrepancies between the Wolf and Group SSN series during the 19th century.

When Rudolf Wolf constructed his Sunspot Number series (SSN) he had the problem of calibrating other observers counts against his own. Since Wolf was one of the discoverers of the relationship between geomagnetic variations and the number of spots it was tempting for him to apply this relationship as a means to securing the calibration. Wolf posited the following linear relationship:

$$rD = a + b\, R_{\mathrm{W}}$$

where rD is the range of the geomagnetic Declination from its extremum in the morning to its extremum in the afternoon and R_{W} is Wolf's newly defined Sunspot 'Relative' Number, namely 10*#Groups + #Spots. Wolf labored the rest of his life to determine the 'constants' a and b, several times lamenting that "by now the last of the doubting Thomases would have to give in and accept my results". As Wolf extended his series each year, he never failed to remark how well his formula held up, even at times [successfully] predicting what value of ***rD*** his correspondents at various geomagnetic observatories would eventually report to him for the year [if the report was late in reaching Wolf].

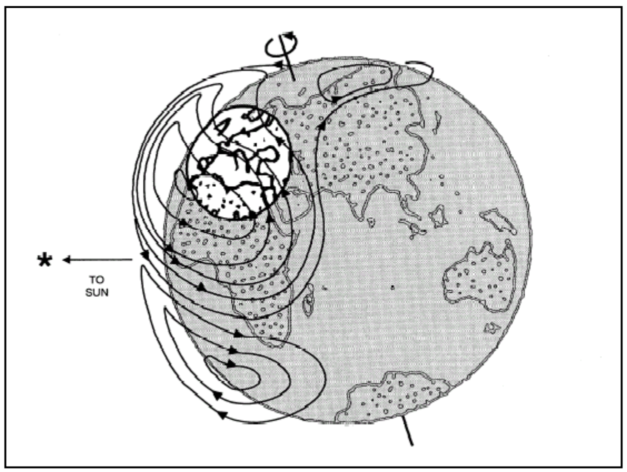
So confident was Wolf in the correctness of his formula that he used it to correct the SSN series not just once, but twice. Between 1857 and 1861, Wolf effectively doubled the counts by Staudacher (purple), and between 1861 and 1875, Wolf increased all values before his own observations in 1849 by 25%:



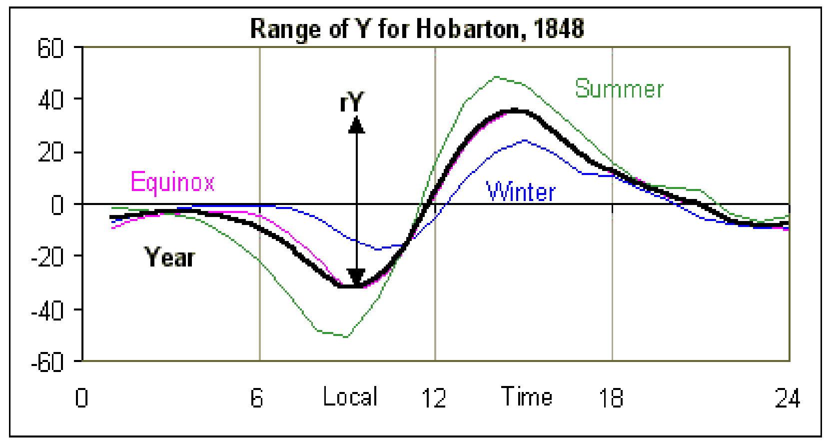
The Figure shows the yearly average Wolf numbers reported by Wolf in a given year (e.g. 1857 - purple crosses) divided by the ‘official’ International (or Zurich) sunspot number given today [omitting years with Rz = 0]. Note that for the Dalton minimum in the early 1800s, the values quoted today are significantly smaller than what Wolf thought in 1880.

Wolf was justified in using geomagnetic data for calibration of the sunspot number

George Graham (1724) discovered in 1723 that the direction (called the Declination today) of the horizontal component of the Earth's magnetic field varies systematically during the day, moving away from its average direction during the morning, then deviating in the other direction during the afternoon [see Figure below]. The movement is slight (a few minutes of arc), but easily measured. The origin of these deflections is the combined magnetic effects of ionospheric current systems flowing in the E-region and of corresponding induced "telluric" currents, created by dynamo action. These systems consist of two vortices, one in each hemisphere, with foci at ~30º latitude and ~1 hour before local noon, both comprising two currents, one flowing above the Earth's surface and the other one (in the opposite direction) underneath the surface. The external current system is shown schematically below. These currents, fixed in space in relation to the Sun, flow at all times, the Earth rotating under them, and give rise to the mostly regular daily variation discovered by Graham, the so-called S_R variation:

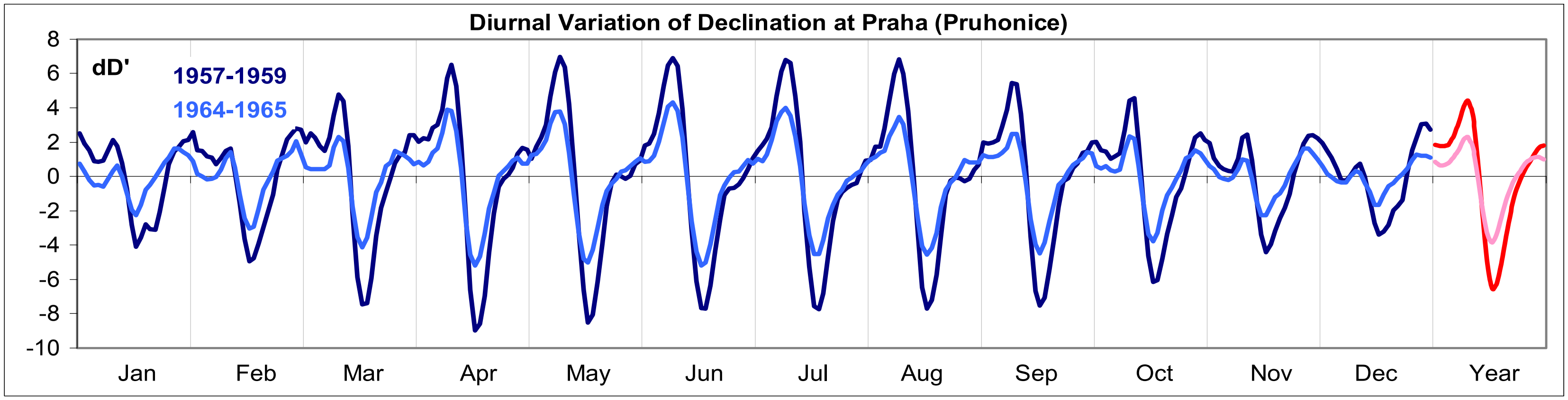


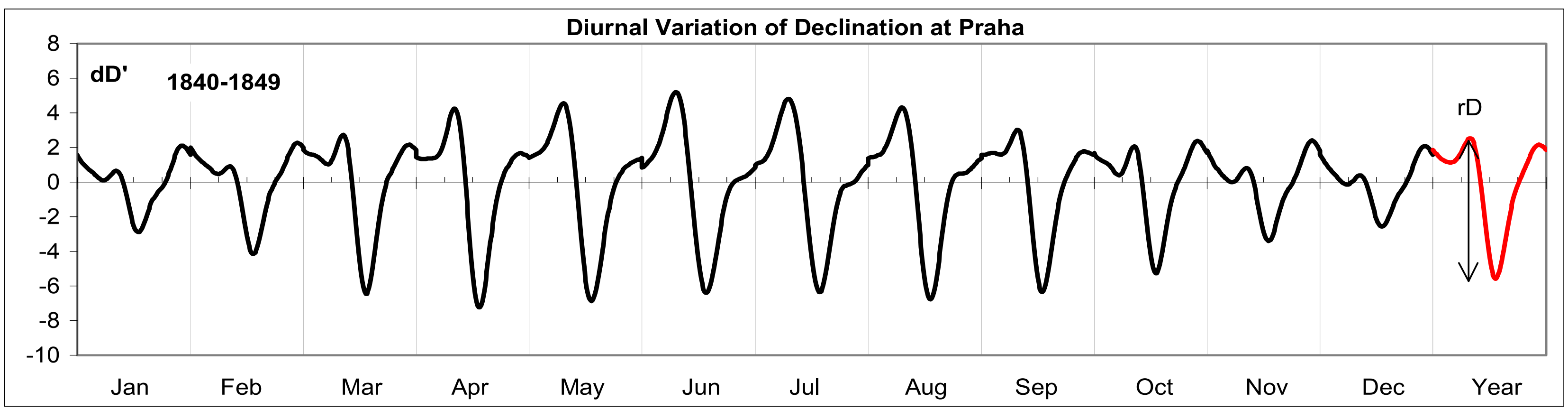
Current system after Torta et al.



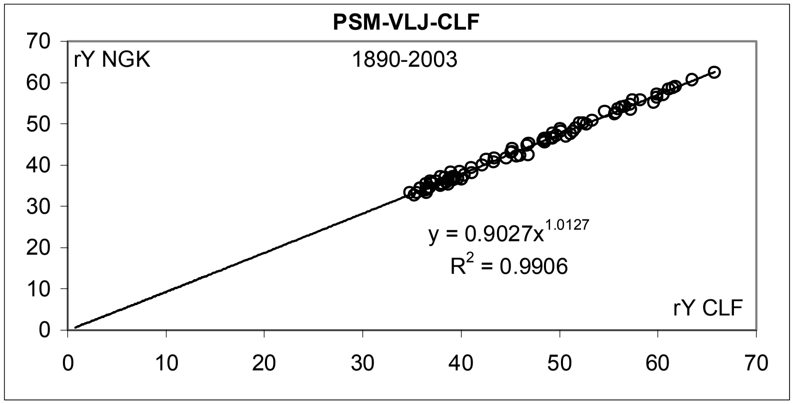
Diurnal variation of Y component at Hobarton, 1848

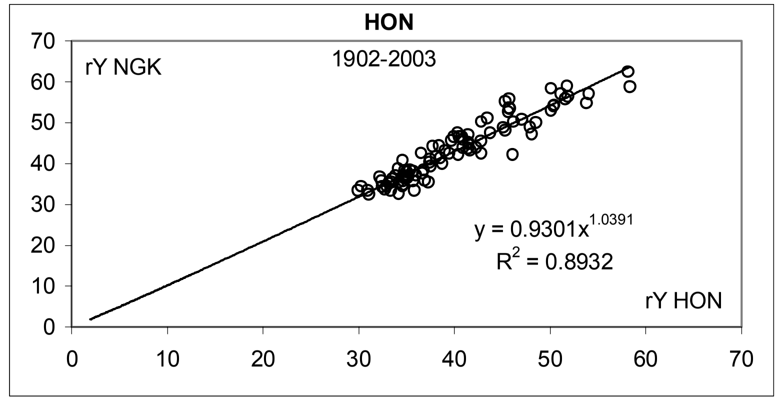
As the current flows North-South along the flank of the current system, the magnetic deflection is in the East-West direction and the correct physical parameter to use is the Y-component of the geomagnetic field, calculated from the horizontal force ***H*** and the Declination ***D*** as $Y = H \sin(D)$ or for the ranges $rY \sim H \cos(D)\ rD$ since $rY \ll H$. This means that the main reason the constants ***a*** and ***b*** in Wolf's formula varied from station to station was that ***H*** varied from station to station. The range in ***D*** was easily measured even in the 1840s as was the seasonal variation:





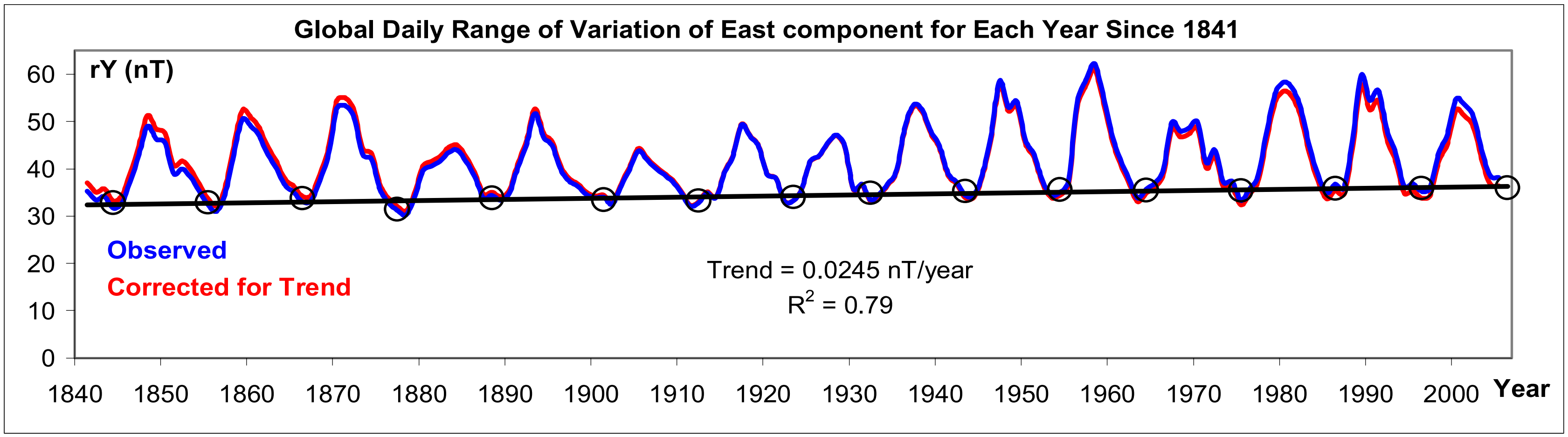
As the internal currents depends on local conductivity of the earth [and sea] different stations will measure a slightly different magnetic deflection. We can compensate for those station differences by normalizing the measured range ***rY*** to a standard station, (Potsdam, Seddin, and) Niemegk near Berlin. This is justified by the very high correlation between stations even on the other side of the Earth (e.g. Honolulu):



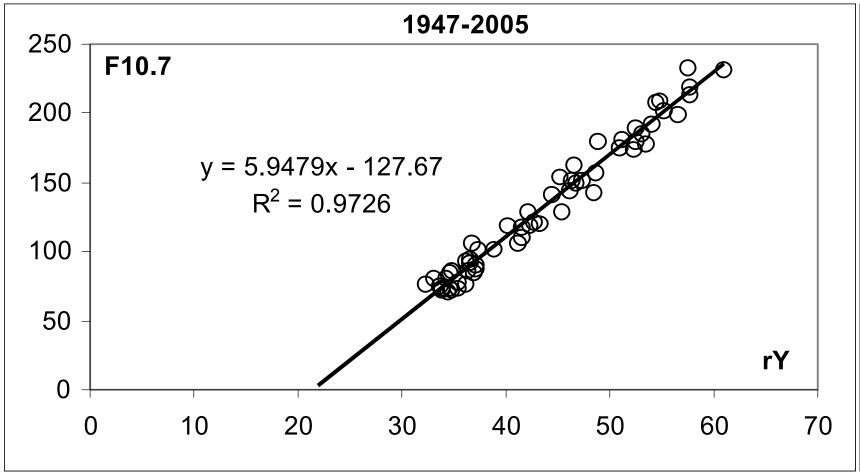


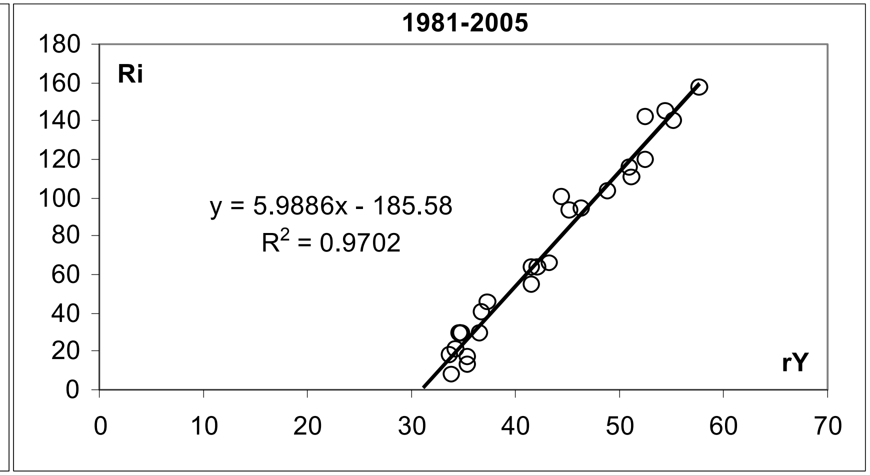
Hobarton-Waterloo-Toolangi	HBT	1841-1854	WAT	1919-1958	TOO	1924-1978
Praha	PRA	1841-1872				
Oslo-Dombas	OSL	1842-1930	DOM	1930-2003		
Helsinki-Nurmijarvi	HLS	1845-1894	NUR	1957-2006		
Batavia	BTV	1868-1928				
Parc-st-Maur-Val Joyeux-Chambon la Foret	PSM	1890-1893	VLJ	1923-1935	CLF	1936-2004
Greenwich-Abinger-Hartland	GRW	1890-1893	ABN	1925-1956	HAD	1957-2006
Potsdam-Seddin-Niemegk	POT	1890-1907	SED	1908-1931	**NGK**	1932-2006
Cheltenham-Frederiksburg	CLH	1902-1956	FRD	1956-2003		
Honolulu	HON	1902-2003				
DeBilt-Witteveen	DBN	1903-1937	WIT	1939-1984		
Vieques-San Juan	VQS	1904-1923	SJG	1925-2004		
Eskdalemuir	ESK	1911-2003				
Kakioka	KAK	1914-2005				
Furstenfeldbruck	FUR	1940-2005				

Normalizing the observatories in the table above to NGK using the reduction formulae given by the correlations (using a third station as 'bridge' if needed) we can construct an unbroken series of ***rY*** values as the median of individual stations for each year since 1841 (blue curve):

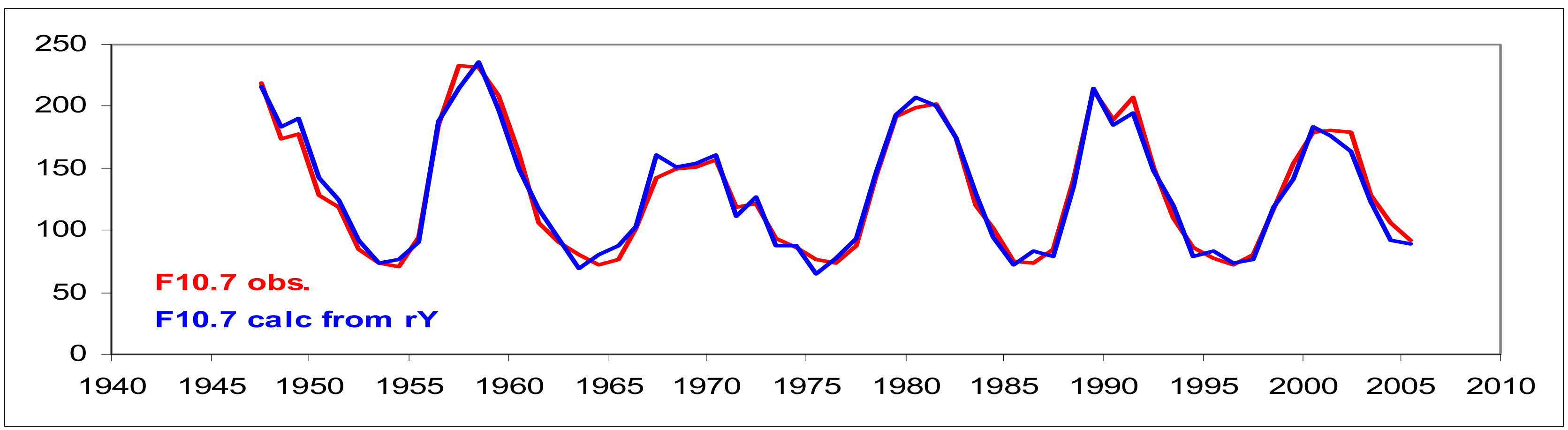


The circles show ***rY*** averaged over the three years around each sunspot minimum. There is a clear trend in these values (0.0245 nT/year) amounting to an increase of 9.8% over the 166-year interval 1841-2006. The red curve shows the ranges with this trend removed. It is likely that the increase simply results from an increase of the ionospheric conductivity caused by the 9% decrease of the Earth's main dipole field over the same time interval. Simple theory predicts that the conductivity should be inversely proportional to the ambient magnetic field strength. The current intensity depends on the ionospheric conductivity (more precisely the height-integrated conductivity over the E-region - the conductance). At low and middle latitudes, the solar FUV in the Schumann-Runge continuum band (between 107 nm and 170 nm) provides most of the ionization. The F10.7 radio flux has been shown to be a good proxy for the FUV flux. We should then expect a good correlation between ***rY*** and the F10.7 flux (available since 1947); this is, in fact, the case with high fidelity (left panel):

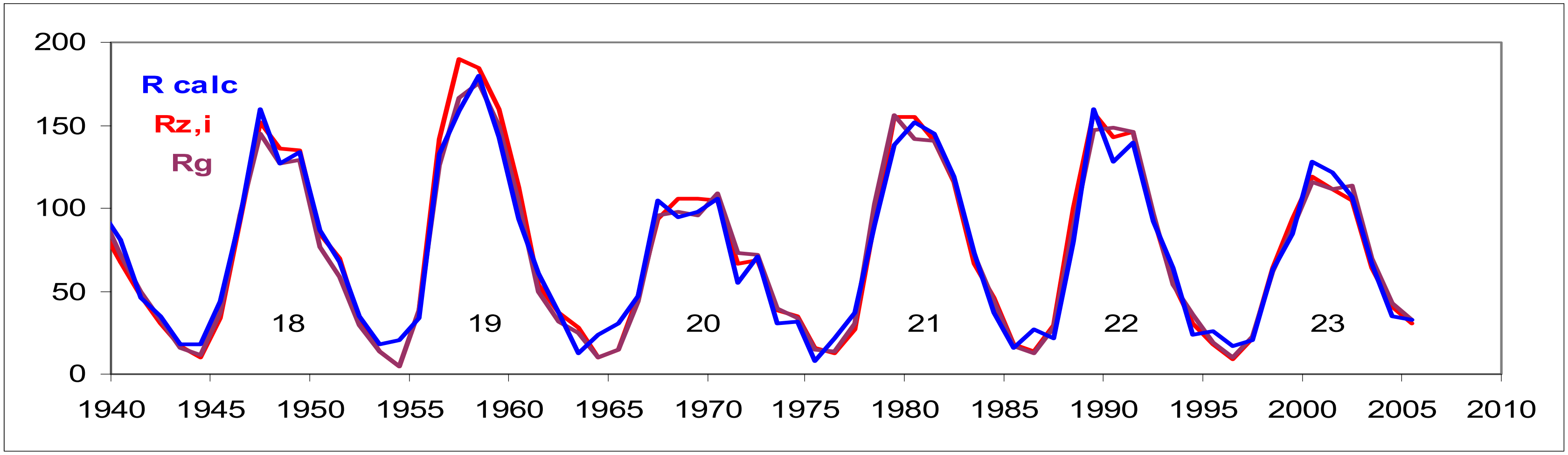




We can now calculate F10.7 from *rY* using the regression established above. In a sense, the good correlation is a validation of the idea that F10.7 is a good proxy for the FUV emission, and is also a validation of the mechanism behind Wolf's relationship using *rY* instead of *rD*.

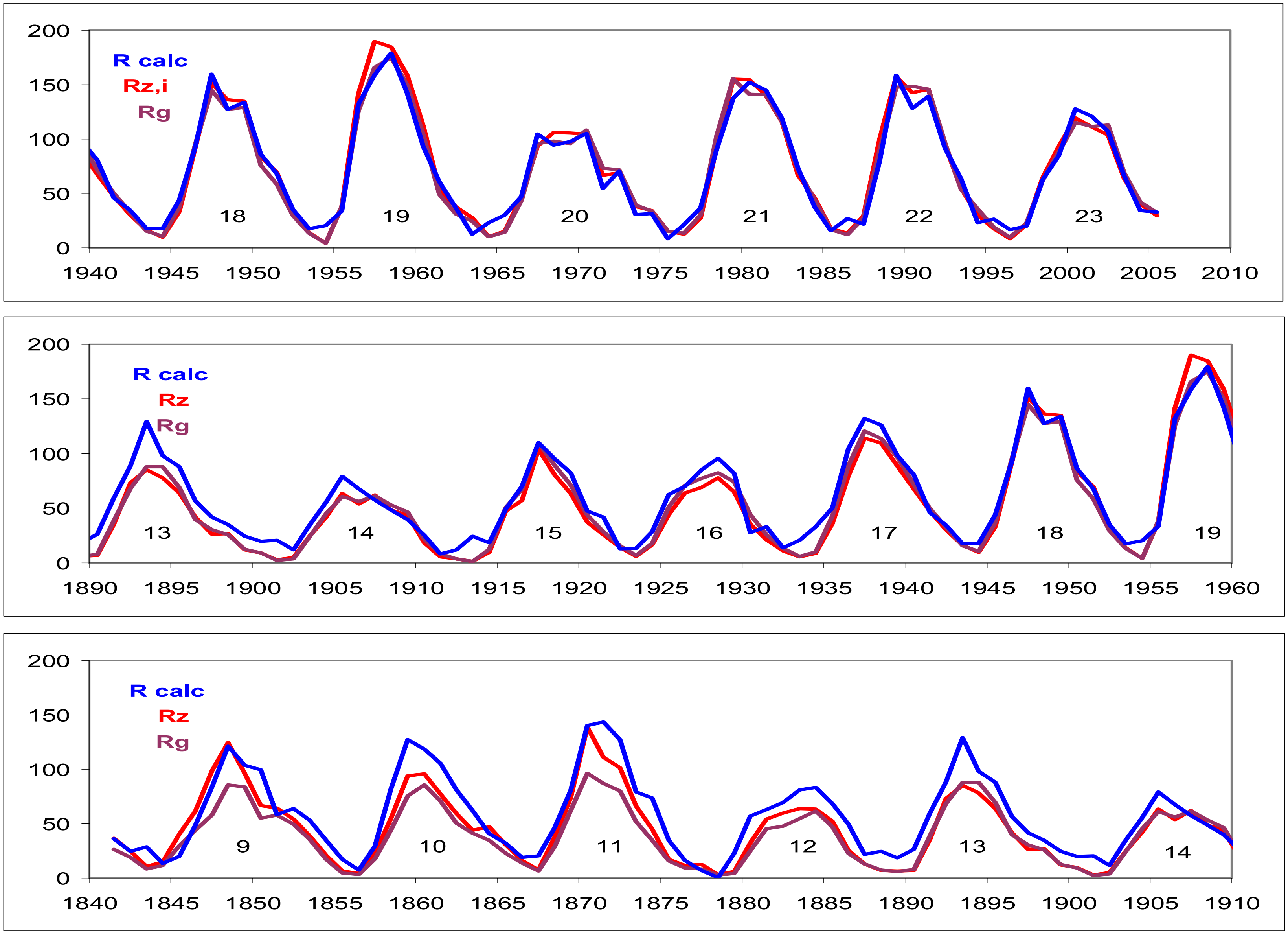


Since there is a good correlation between F10.7 and the sunspot number we expect [and find] a good correlation between $\boldsymbol{rY}$ and the international sunspot number, R_I, which after 1980 is assumed to have uniform definition and calibration. Using then the regression equation for R_I, we calculate R_I from $\boldsymbol{rY}$. The result (blue curve) is shown below back to 1940. Also shown are yearly averages of observed R_I, or R_Z (Zurich, red) and R_G (Group, purple).



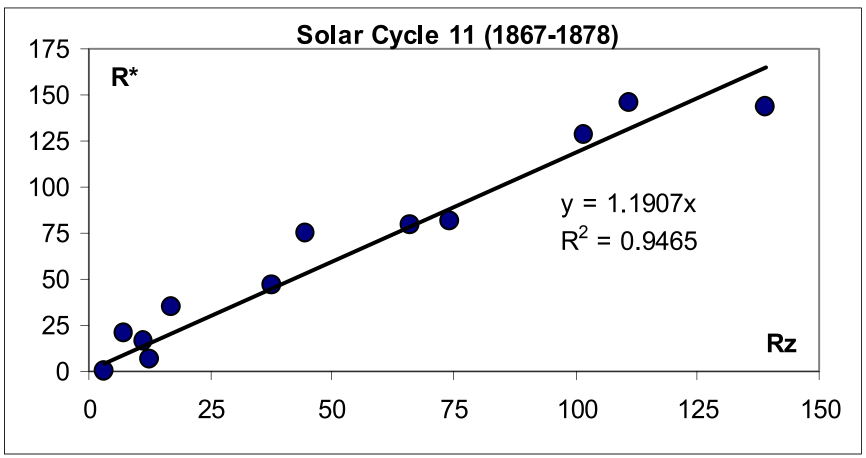
During these last 65 years, all the activity measures agree reasonably well, with the possible exception of the maximum of cycle #19, which is somewhat too high compared to the Group number and the sunspot number calculated from the measured $\boldsymbol{rY}$.

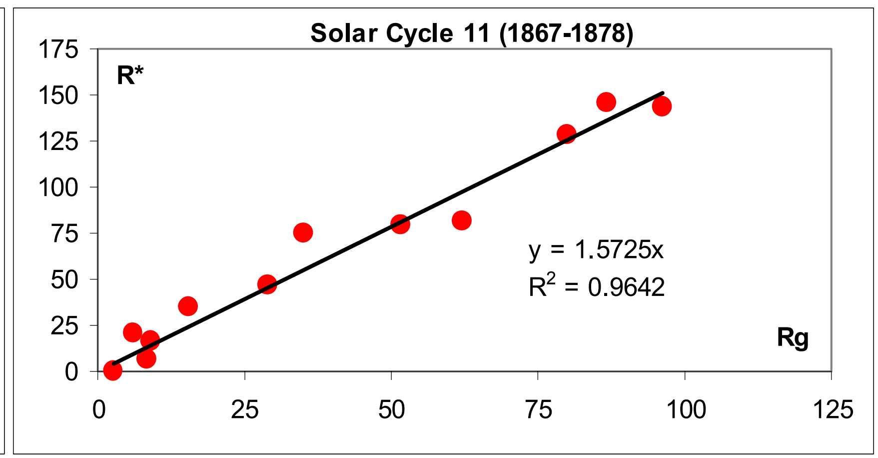
We'll now make the crucial assumption, namely that relationship established during modern times between the sunspot number and the diurnal range of the $\boldsymbol{Y}$-component holds in the past as well, and calculate the sunspot number from the daily range, $\boldsymbol{rY}$, using the regression equation valid for 1981 and later. The following Figure compares the calculated sunspot number (blue curve) with the official "international" sunspot number (red curve) and the "Group" sunspot number:


R calc
Rz,i
Rg
18
19
20
21
22
23
0
50
100
150
200
1940
1945
1950
1955
1960
1965
1970
1975
1980
1985
1990
1995
2000
2005
2010
R calc
Rz
Rg
13
14
15
16
17
18
19
1890
1895
1900
1905
1910
1915
1920
1925
1930
1935
1940
1945
1950
1955
1960
R calc
Rz
Rg
9
10
11
12
13
14
1840
1845
1850
1855
1860
1865
1870
1875
1880
1885
1890
1895
1900
1905
1910

It is evident that the observed sunspot numbers generally match our reconstructed sunspot numbers back to the mid-1940s, but that the observed sunspot numbers generally fall below our reconstruction before that; the difference increasing as we go further back in time. The difference is largest for the Group sunspot numbers. Occasionally, Wolf got it 'right', *e.g.* for cycle 9 with maximum in 1848. The differences are at times very large, up to 50%.

It is clear that the differences between the calculated sunspot number and the two observed series vary with time. No single trend is apparent, so we opt for finding a correction factor separately for each cycle by fitting the reconstructed and observed values by a straight line through the origin as shown below for cycle 11. We thus de-emphasize the influence of just the maximum value and spread the correction evenly (in the least-squares sense) over the entire cycle. The slopes of the trend lines give the correction factor in each case:

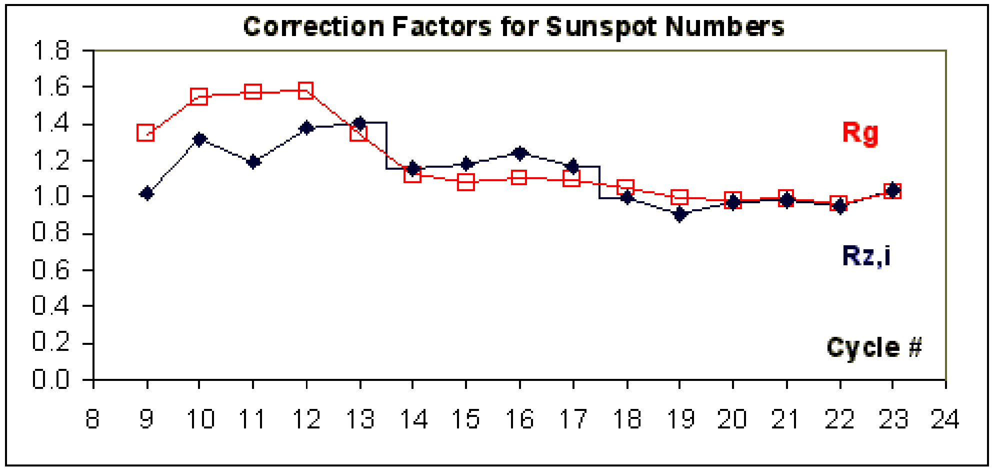




We now construct the following Table with the correction factors for each series, R_Z and R_G, for each cycle to be applied to each year (stipulating the same factor for monthly and daily values) within the cycle. That is, we assume that the calibration is constant within a cycle. This can, at best, only be an approximation to the truth, but can be justified by the finding that none of the correlation plots for any of the other cycles show any clear jumps or other signs of a mixture of two populations with different calibration.

Cycle	Rg	Rz
9	1.347	1.020
10	1.545	1.317
11	1.572	1.191
12	1.580	1.379
13	1.343	1.403
14	1.121	1.156
15	1.075	1.180
16	1.103	1.238
17	1.093	1.166
18	1.046	0.996
19	0.991	0.905
20	0.978	0.970
21	0.986	0.982
22	0.961	0.948
23	1.026	1.036

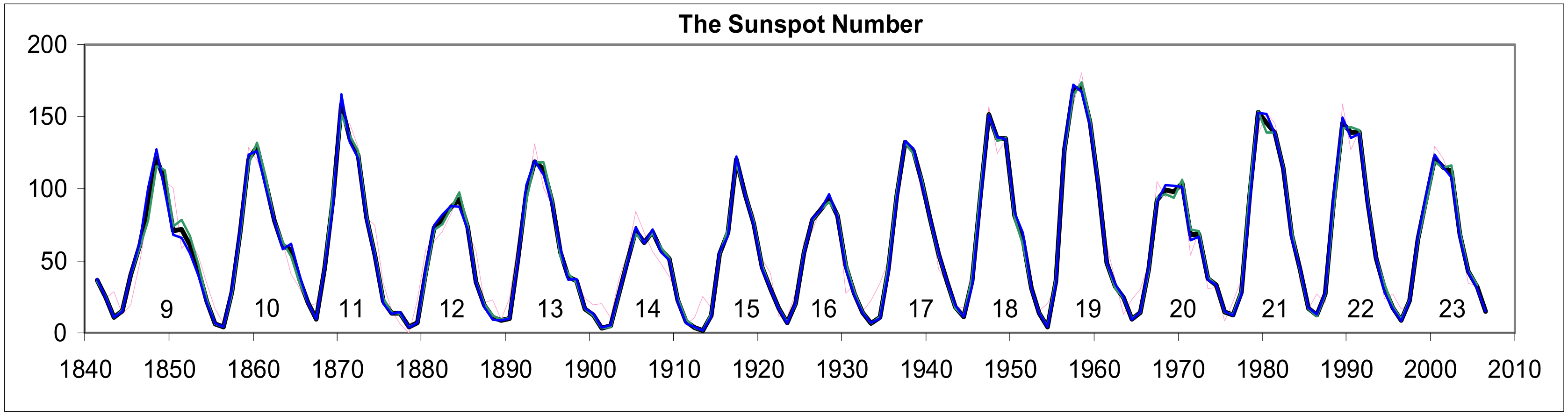
Correction Factors for reported sunspot numbers.



Plot of Correction Factors to apply to R_Z and R_G for each cycle

The counting method for small spots changed in 1893 when Wolf died and his assistant Wolfer carried on the series, and apparently in 1945 when Waldmeier took over. These changes seem to be duly reflected as discontinuities in the inferred correction factor for the Zürich sunspot number as shown above.

We can now plot the corrected sunspot number series since cycle 9. There is no real difference between the corrected Group sunspot numbers and Zürich sunspot numbers. Both are plotted, but the curves fall on top of another. It is of interest to note that (corrected) cycles 11 and 10 were as active as the most recent cycles 22 and 23. We thus see no evidence in the sunspot number of a secular increase in solar activity over the last ~165 years:



Conclusion:

Already the fact that the Zürich and the Group sunspot numbers are different before ~1875 should give pause. That neither of them is consistent with the observed variation of the daily range of the geomagnetic S_R variation might be a hint that the debate is not which of the two to use, but how to reconcile the observations into a consistent dataset. We suggest that careful analysis of geomagnetic data (extending back into the 1740s) could be a possible approach to securing the calibration of solar activity over time, which has taken on a new importance as an element in the debate over climate change.