Calibrating the Sunspot Number using "the Magnetic Needle"

Leif Svalgaard

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. 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 in Figure 1. 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.

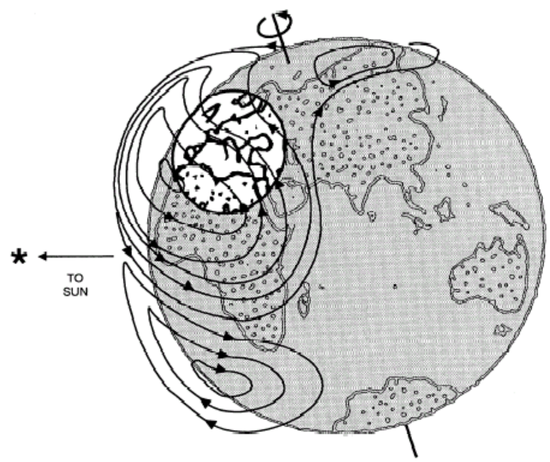


Figure 1. Schematic representation of the S_R ionospheric current system (after Torta *et al.*, 1997).

The current intensity and the size of the vortices change with the seasons, being largest in the summer hemisphere. In addition, there are variations of the position of the foci and of the current intensity (by up to factor of two) from day to day, so the magnetic effects at a given station can show quite complex patterns. Along the 'flanks' of the (external) vortices, the current flow is equatorwards on the morning side and polewards on the afternoon side. The magnetic effect at mid-latitudes of these currents at a right angle to the current flow are thus East-West and rather insensitive to the position of the focus. As the magnetospheric ring current and the auroral electrojets and their return currents that are responsible for geomagnetic activity have generally North-South directed magnetic effects (strongest at night), the daytime variation of the Y or East component (Y positive towards East, X towards North, and Z downwards) is a suitable proxy for the strength of the S_R ionospheric current system.

Figure 2 shows the daily variation of the East component of the geomagnetic field at Hobarton (Tasmania, 42.9º lat. South) for the year 1848. We can define the range rY (in nT) as the yearly average peak to valley difference. Note the seasonal variation of the amplitude of the diurnal variation. By averaging over a year, the seasonal variation evens out and minor season-dependent shifts and irregularities are minimized.

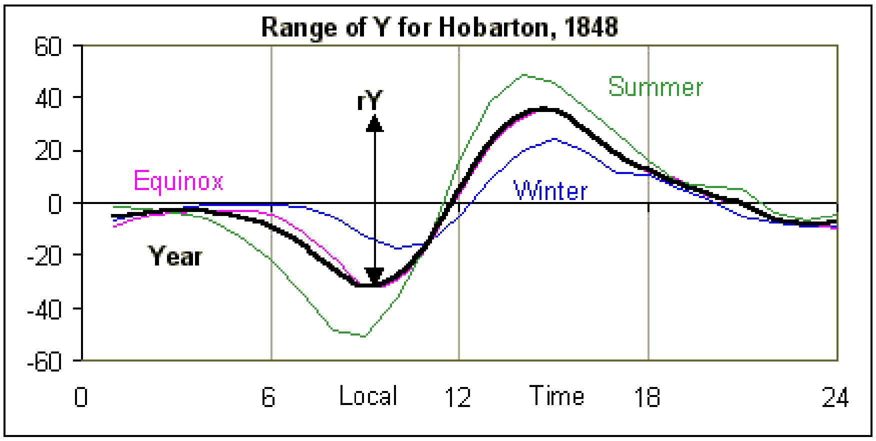


Figure 2. Daily variation of the East component (nT) of the geomagnetic field at Hobarton for the year 1848.

We have reliable measurements like the one shown above for many observatories (the number ranging from a handful in the 1840s to more than a hundred in the 21st century). Selecting mid-latitude stations, one finds that the diurnal range, rY, for each year does not vary much (less than a factor of two) from station to station. Using overlapping data we normalize the yearly values of rY for each station to that of the Niemegk station (NGK, in Germany) and plot the average 'global' range as a function of time in Figure 3.

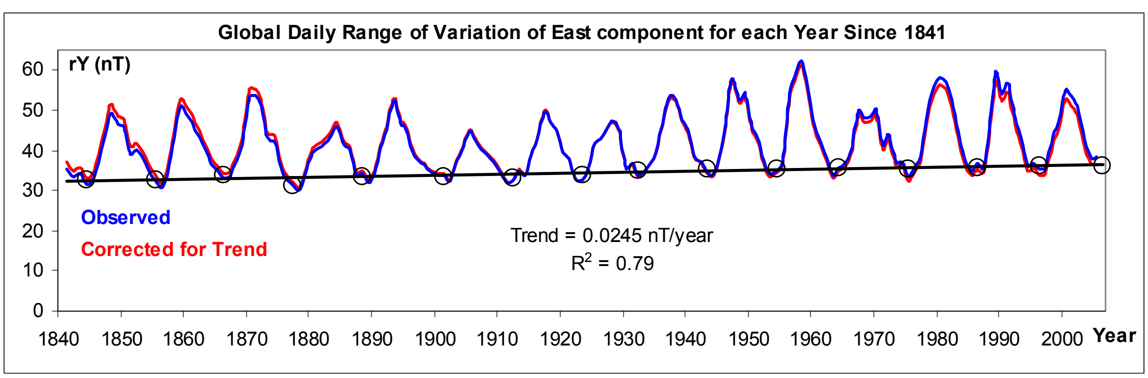


Figure 3. The average yearly range, rY, of the daily variation of the East component of the horizontal force (blue), corrected for secular trend (red).

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.

Of special interest to us for this article is the clear solar cycle variation of rY. This was noted already by the earliest observers of the sunspot cycle and geomagnetic variations (Sabine, Wolf, Gauthier, Lamont) around 1850. Rudolf Wolf codified the relationship by a linear relation $rD = a + b\ R_W$, where rD was the range of the Declination measured in minutes of arc (rather than the range of the Y-component in nT) and R_W was the 'Wolf' number given by $R_W = k(10\ g + f)$, where g was the number of sunspot groups, f the number of spots, and k a calibration factor. Wolf observed the sunspots from 1849 until his death in 1893. By 1861, he was able to publish yearly sunspot numbers from 1749 through 1860 (Wolf, 1861). He noted what he considered to be "one of the most important relations among the Solar actions yet discovered" namely that "greater activity on the Sun goes with shorter periods, and less with longer periods". In the following years, Wolf also collected data on rD from observatories all over the world and became more and more convinced of the basic validity of his linear relation. In fact, in his yearly reports on the sunspot number he never failed to compare the observed yearly values of rD with the value calculated from his relationship and always found "good agreement", thus validating the sunspot number against an independent observable. So strong had the confirmation become that Wolf around 1880 quietly revised his sunspot series by increasing the pre-1849 values by ~20% to bring them into better agreement with the geomagnetic record.

A valid criticism of the use of rD is that it is the result of the pull of two force vectors, the (nearly constant during the day) largely North-South horizontal force of the main geomagnetic field and the (varying during the day) largely East-West force of the magnetic effect of the S_R current system., and that therefore the range of the angle in arc minutes varies with the horizontal force as well, in space and in time. François Arago wonderfully described (in the 1820s) how the range of the Declination he observed at the Paris Observatory increased by a factor of ten as the result of installation (later removed) of an iron stove in an adjoining room, the magnetic stove canceling out a part of the natural horizontal force. Later researchers did not share Wolf's enthusiasm for his relationship. With our modern understanding we realize that the relation is sound except that the proper variable to use is the East-West deflection, rY in force-units (nT), having a direct physical interpretation in terms of current intensity.

The current intensity in turn 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 (between 107 nm and 170 nm) band provides most of the ionization. The F10.7 radio flux has been shown (*e.g.* Strobel, 1978) 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). Figure 4 shows this to be the case with high fidelity.

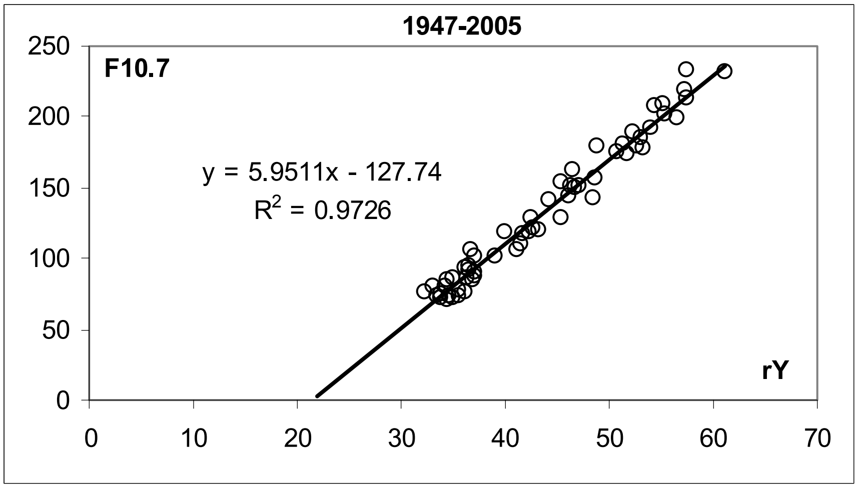


Figure 4, Relation between yearly average solar F10.7 radio flux (solar flux units) and the average East component range, rY (nT) for the interval 1947-2005.

We can now calculate F10.7 (Figure 5) 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.

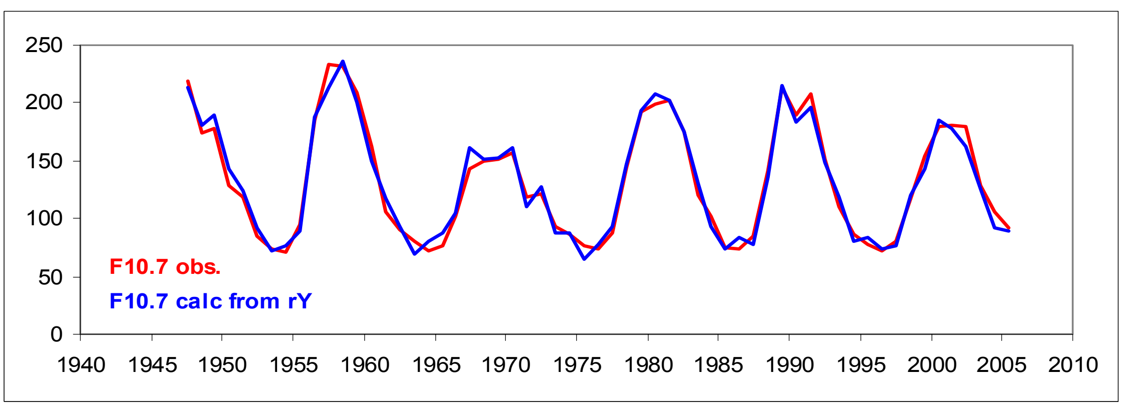


Figure 5. Yearly averages of observed (red curve) F10.7 radio flux (solar flux units) and calculated (blue curve) for 1947-2005. In this and several of the other Figures, the differently colored curves will often overlay each other signifying near perfect match.

We can thus reconstruct the F10.7 radio flux as far back as we have rY data (to 1740 with some gaps). Because there is also a good correlation between the F10.7 radio flux and the sunspot number, we expect a good correlation between rY and $R\beta$, β might be any of I

(International from 1981), Z (Zürich, 1749-1980), or G (Group, 1610-1995). Assuming that the international sunspot number, R_I, after 1980 has a uniform definition and calibration, we can check the expectation (Figure 6).

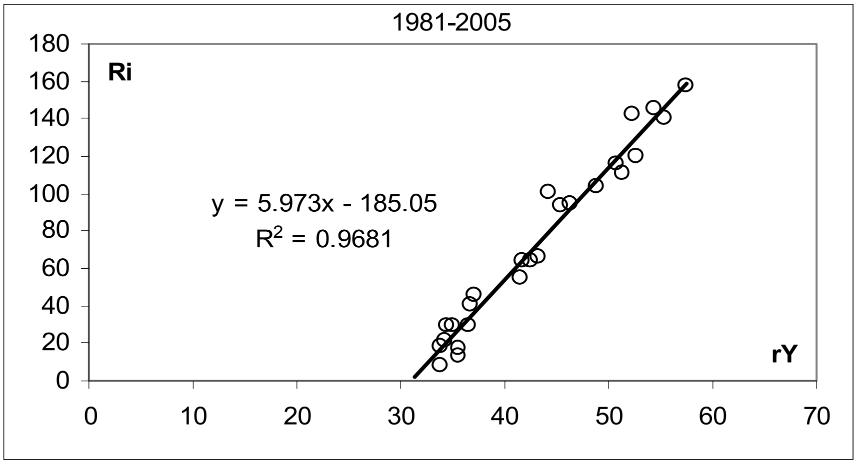


Figure 6. Relation between yearly average International sunspot numbers, R_I, and the average East component range, rY (nT) for the interval 1981-2005.

The correlation is just as good as for F10.7, 97% of the variation matches that of rY; rather remarkable, considering the arbitrariness of the sunspot number definition. If one correlates other subsets of the sunspot series, one finds similar correlations, but the regression constants are different. This suggests that the sunspot series does not have calibration that is constant in time, assuming that our understanding and analysis of rY is correct. As the definition of the sunspot number is arbitrary and the counting process somewhat subjective, there is no 'correct' sunspot number. It seems most practical to adopt the International Sunspot Numbers since 1981 as the base for any standardization, in other words to adopt the regression equation of Figure 6. It is generally accepted that the relation between the sunspot number and the F10.7 flux is weakly quadratic with a slight upturn for sunspot numbers less than ~15. For the 1981-2005 interval, the quadratic term is very small and not statistically significant. The main reason for the upturn may be that the definition of the sunspot number has no values between 0 and $11k$. For the calibration of the sunspot number these very low values have no great import so we shall use the simple linear regression equations.

Using then the regression equation for R_I, we calculate R_I from rY since 1841. The result (blue curve) is shown in Figure 7. Also shown are yearly averages of observed R_I, or R_Z (red) and R_G (grey).

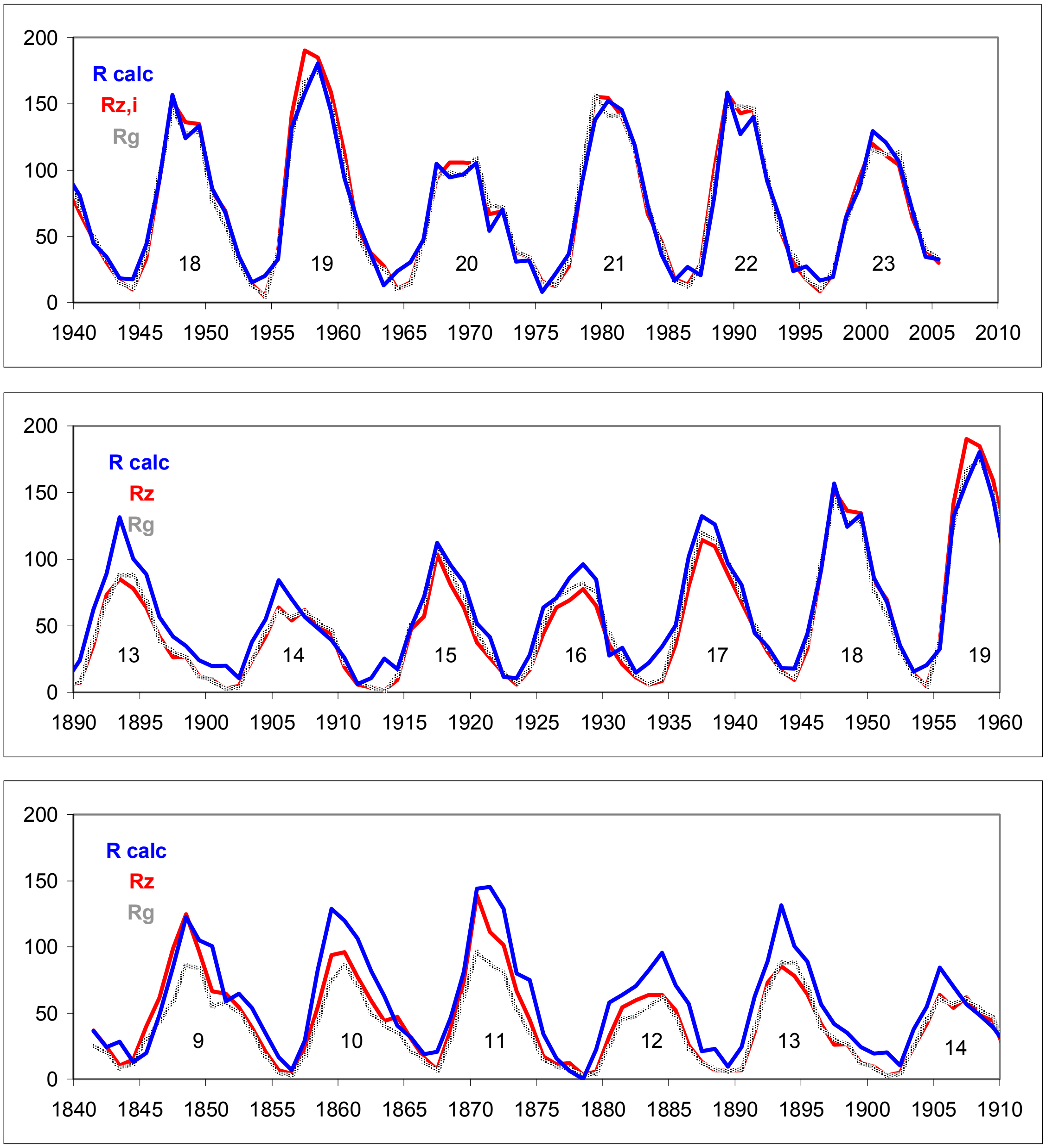
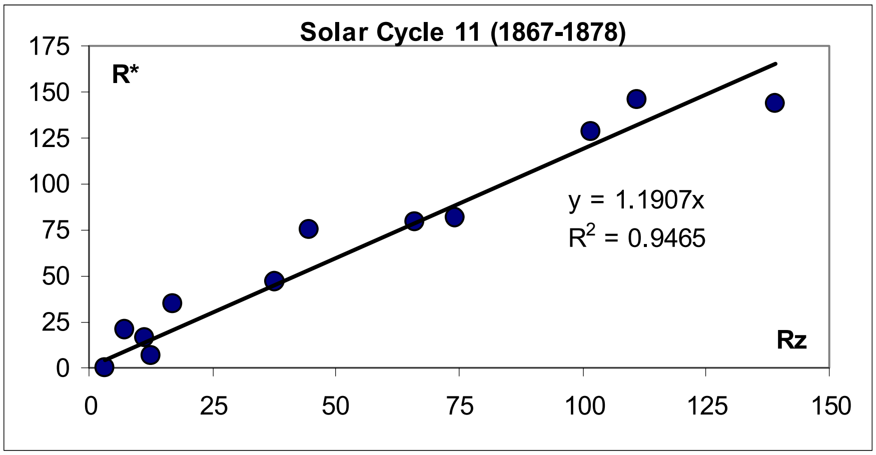


Figure 7. Calculated (blue curve) of yearly average R_I from rY since 1841. Observed yearly averages of R_I, or R_Z (red) and R_G (grey) are shown for comparison. Each cycle is marked with the number of the cycle. Note the overlap between panels.

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 in Figure 8 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.



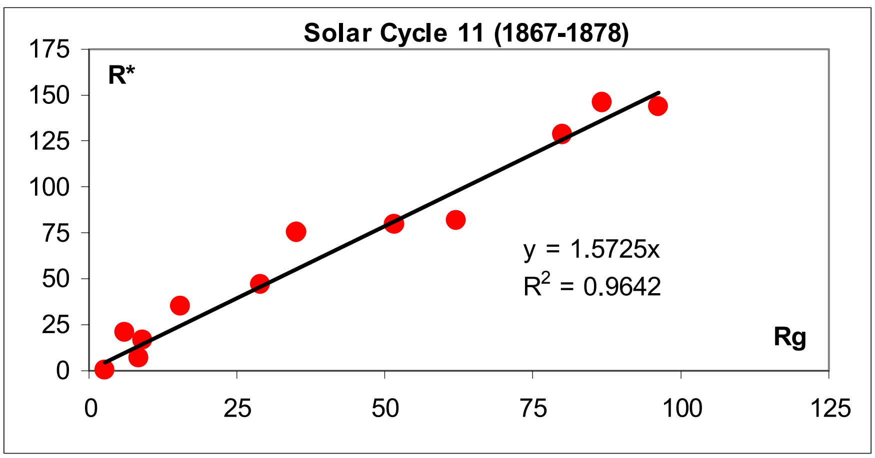


Figure 8. (Top) Correlation between calculated R_I and observed R_Z for each year of cycle 11 forced to pass through the origin. (Bottom) Correlation between calculated R_I and observed R_G for each year of cycle 11 forced to pass through the origin. The slopes of the trend lines give the correction 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 like Figure 8 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.

The counting method for small spots changed in 1893 when Wolf died and his assistant Wolfer carried on the series, and 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 in Figure 9.

It is not clear why the Group Sunspot Number calibration changes. If we were to entertain the view that R_G is so simple to measure that it has to be correct, we must accept that rY must either be in error by more than 50% or that our interpretation of the cause of rY is seriously incomplete. It must be remembered that rY is also easy to measure. Wolf was correct in insisting that the geomagnetic effects should track the sunspot numbers according to some relation with a physical interpretation.

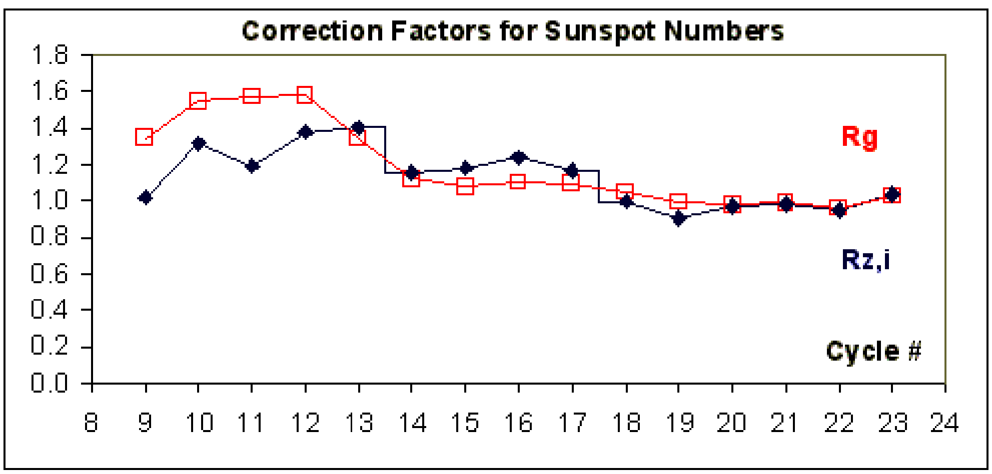


Figure 9.Factors for each cycle to scale the Zürich sunspot number (blue) to match R_I calculated from rY (in turn scaled to R_I from 1981 on) and to scale the Group sunspot number (red) to calculated R_I .

We can now plot (Figure 10) the corrected sunspot number series since cycle 9. There is no real difference between the corrected Group sunspot numbers and Zurich 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.

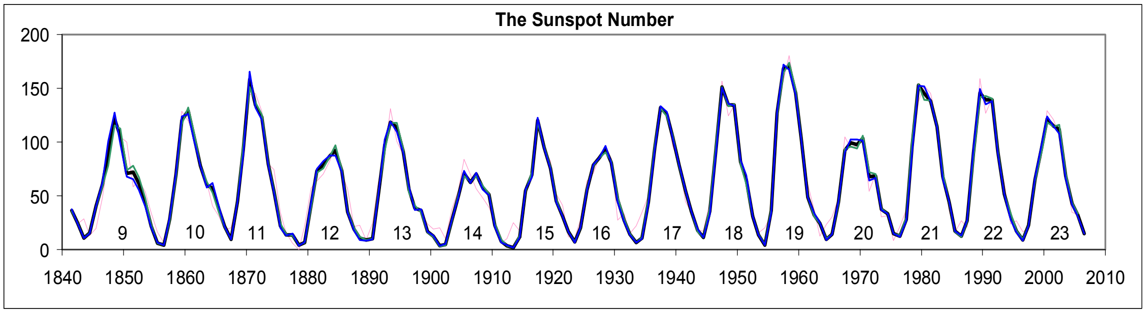


Figure 10. Corrected yearly average sunspot numbers 1841-2006 consistent with the long-term variation of the daily range of the geomagnetic East-component.

Conclusion: What is one to make of this? 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.

References

Graham, G. (1724) An Account of Observations Made of the Variation of the Horizontal Needle at London, in the Latter Part of the Year 1722, and Beginning of 1723, *Phil. Trans. Roy. Soc., 33*, 96-107, doi: 10.1098/rstl.1724.0020.

Strobel, D. F. (1978) Parameterization of the atmospheric heating rate from 15 to 120 km due to O2 and O3 absorption of solar radiation, *J. Geophys. Res., 83*, 6225-6230.

Torta, J. M., J. J. Curto, & P. Bencze (1997) Behavior of the quiet day ionospheric current system in the European region, *J. Geophys. Res.*, *102 (A2)*, 2483-2494.

Wolf, R. (1861) Abstract of his latest Results, *M. Notices Roy. Astron. Soc.*, *21*, 77.