

Does Correction Factor Vary with Solar Cycle?

Heon-Young Chang^{1†} and Sung-Jin Oh²

¹Department of Astronomy and Atmospheric Sciences, Kyungpook National University, Daegu 702-701, Korea

²Butterstar Observatory, Dongduchen High School, Dongduchen 483-707, Korea

Monitoring sunspots consistently is the most basic step required to study various aspects of solar activity. To achieve this goal, the observers must regularly calculate their own correction factor k and keep it stable. Relatively recently, two observing teams in South Korea have presented interesting papers which claim that revisions that take the yearly-basis k into account lead to a better agreement with the international relative sunspot number R_i , and that yearly k apparently varies with the solar cycle. In this paper, using artificial data sets we have modeled the sunspot numbers as a superposition of random noise and a slowly varying background function, and attempted to investigate whether the variation in the correction factor is coupled with the solar cycle. Regardless of the statistical distributions of the random noise, we have found the correction factor increases as sunspot numbers increase, as claimed in the reports mentioned above. The degree of dependence of correction factor k on the sunspot number is subject to the signal-to-noise ratio. Therefore, we conclude that apparent dependence of the value of the correction factor k on the phase of the solar cycle is not due to a physical property, but a statistical property of the data.

Keywords: Sun, sunspot, data analysis

1. INTRODUCTION

The sunspot number provides the longest available record of solar activity. Although its short-term variations give some important insight, e.g. regarding solar differential rotation, its long-term behavior and the longevity of its series may provide many more such insights. This is why the international scientific community has continuously renewed its interest in the sunspot number. Many scientific investigations, including cycle analysis and forecast, solar North-South asymmetry analysis, coherence analysis of the solar magnetic field, mid-term studies of solar activity, are based on the Solar Influences Data Analysis Center (SIDC) sunspot data (Pulkkinen et al. 1999, Lockwood 2003, Solanki et al. 2004, Wang 2004, Chang 2007, Usoskin 2008, Petrovay 2010, Ternullo 2010, Kim & Chang 2011).

The SIDC (Berghmans et al. 2005) collects monthly observations from various stations worldwide in order to calculate the International Relative Sunspot Number R_i (Letfus 2000, Clette et al. 2007, Vaquero 2007). The

center broadcasts the daily, monthly, and yearly sunspot numbers, with middle-range predictions (up to 12 months). The Sunspot Index Data Center was founded in 1981 to continue the work, when the Zürich Observatory decided to halt computing and publishing the sunspot number. In 1981, the Sunspot Index Data Center began producing a sunspot index, called the International Relative Sunspot Number, R_i . Continuity and coherence with the former index of Zürich was assured through the use of Locarno as a reference station. In 2000, reorganized and expanded by gaining the qualification of Regional Warning Center of the International Space Environment Service, the Sunspot Index Data Center had become the SIDC. The bulletins and reports of the SIDC are all freely available for the scientific community and the general public on the internet (<http://sidc.oma.be>).

In 1849, Wolf at the Zürich Observatory proposed the widely used formula: $R_z = 10g + f$, where f is the number of individual sunspots, and g is the number of sunspot groups (Letfus 2000, Clette et al. 2007, Vaquero 2007). The

© This is an open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Received Apr 23, 2012 Revised May 7, 2012 Accepted May 9, 2012

†Corresponding Author

E-mail: hyc@knu.ac.kr

Tel: +82-53-950-6367 Fax: +82-53-950-6359

quality factor k was introduced later in order to compare results from different observers, the stability of the Earth's atmosphere and telescopes, giving the formula $R_i = k(10g + f)$. Observers must satisfy several criteria to be included in the SIDC network: dedication (more than 10 observations per month), regularity (no missing months) and consistency. It is the last that is quantified through the quality factor k , which is the correction factor between the raw sunspot number of an individual station and the global network average. The correction factor k , which typically has a value between 0.4 and 1.7, ensures that results can be compared with each other. The coefficient of the reference station Locarno is fixed at the value of $k_{Loc} = 0.6$. The task of the SIDC consists of collecting the observations from as many stations as possible worldwide, determining the appropriate k factor for each of them, and extracting an overall R_i from all these observations in a good statistical sense.

Let us here briefly describe the statistical processing of input data coming from the worldwide observing network (~90 stations located in ~30 different countries). The R_i processing consists of two main steps. In the first step, the daily reduction coefficients to Locarno are calculated and monthly averaged for every station. For each station, daily values deviating by more than 2σ from the monthly station average are eliminated. The monthly averages are recomputed iteratively until k values are consistent for all stations. In the second step, using the updated monthly k coefficients, the R_i value is computed for each station, and network averages R_d and standard deviations σ are computed for each day. Elimination on the basis of a 1-sigma criterion is used on newly calculated daily means, until the number of retained stations remains unchanged, or the final relative standard deviation is lower than 10%. The final result is retained as the daily R_i . The final quality control consists essentially of regular comparisons between the sunspot number R_i on one hand, and an average of about 20 selected good stations (including the Locarno reference station) or the 10.7 cm radio flux on the other.

Recently, Oh & Chang (2012) have reported results of sunspot observations at the ButterStar observatory for 3364 days, from the 16th of October in 2002 to the 31st of December in 2011. By applying the linear least-squares method between the observed sunspot number (R_B) and the International Relative Sunspot Number (R_i), the overall correction factor k_b for the entire observing period was found to be 0.9519 with a standard deviation of 0.006. In addition, they attempted the same procedures in each year from 2002 to 2011. When calculated for each year, the yearly correction factor has slightly different k_b values from year

to year, and furthermore shows a trend of changing along the solar cycle. That is, the yearly correction factor k_b is larger during the solar maxima and smaller during the solar minima, in general. It was then considered that it seems possible to reduce the errors in indexing sunspot numbers a little bit further by determining the correction factors year by year. Interestingly, similar conclusions were drawn by Kim et al. (2003), in which the observed data at the Korea Astronomy and Space Science Institute (KASI) were analyzed.

In this paper, we attempt to tackle the question of whether the variation in the correction factor is really coupled with the solar cycle. We also investigate the question of whether this apparent dependency is due to a possibility of lower statistical confidence in the linear least-squares method with a decreasing number of sunspots. If the apparent dependency is indeed due to an artifact in the statistical treatment, researchers should refrain from correcting the observed sunspot numbers R_B on a yearly basis, even though it seems to result in a better agreement. This paper is organized as follows. In Section 2, we briefly introduce the artificial data set used in the present analysis. The results we have obtained are presented in Section 3. Finally, we discuss our results and make some conclusions in Section 4.

2. ARTIFICIAL DATA

We model the sunspot number with random noise superposed on a slowly varying background to characterize the sunspot numbers, as per Chang (2008). The underlying part of the sunspot number data is assumed to represent a sum of undamped oscillators. Thus, the international relative sunspot number R_i is assumed to be approximated by

$$R_i = (1 + \varepsilon) \times \sin^n(\omega_0 t + \phi), \quad (1)$$

where ω_0 represents the solar cycle frequency, ϕ is the phase shift, and ε is random noise. The value of n can be chosen arbitrarily as long as the resulting function resembles the observed sunspot number data. Throughout the paper, ω_0 corresponds to 11 years. We employ two kinds of noise distributions for ε in the current paper: 1) uniform distribution, with which random numbers are distributed uniformly between 0 and 1, and 2) exponential distribution with a unit mean and deviation. It is observed that the multiplicative random noise reproduces observational features more satisfactorily (Chang 2008). This is why we adopt the multiplicative random noise rather than the additive random noise.

To simulate calculating the correction factor k , we need to generate a time series of the observed sunspot number R_B given by

$$\begin{aligned} R_B &= R_i + \alpha \varepsilon_{\text{uni}} \times \sin^n(\omega_0 t + \phi) \\ &= (1 + \varepsilon + \alpha \varepsilon_{\text{uni}}) \times \sin^n(\omega_0 t + \phi), \end{aligned} \quad (2)$$

where ε_{uni} represents the random error which is assumed to obey the uniform distribution between 0 and 1, and α is a measure of signal-to-noise ratio. Other symbols have the same meanings as in Eq. (1). We consider ε_{uni} occurs due to different degrees of the skill of various observers, the atmospheric stability of the observing site, the capacity of telescopes, and so on. One may also consider α as

$$\begin{aligned} R_B &= R_i + \alpha \varepsilon_{\text{uni}} R_i \\ &= (1 + \varepsilon)(1 + \alpha \varepsilon_{\text{uni}}) \times \sin^n(\omega_0 t + \phi) \\ &= (1 + \varepsilon + \alpha \varepsilon_{\text{uni}} + \alpha \varepsilon_{\text{uni}} \varepsilon) \times \sin^n(\omega_0 t + \phi), \end{aligned} \quad (3)$$

whose only difference is the negligible term, $\alpha \varepsilon_{\text{uni}} \varepsilon$. We take this particular form of multiplicative noise since it agrees with the observed features quite well (Chang 2008). That is, in the period with a large number of sunspots, uncertainties are relatively large due to the large number of sunspots.

3. RESULTS

In Fig. 1, we show the monthly average of modeled sunspot numbers generated with the exponentially distributed random noise, ε , as a function of time in a month. The abscissa is time in months elapsed since a solar cycle begins, and the ordinate is measured in an arbitrary unit. The solid line represents R_i generated by Eq. (1). The dotted and the dashed lines correspond to R_B generated by Eq. (2) with $\alpha = 0.1$ and $\alpha = 1.0$, respectively. In Fig. 2, similarly, we show the monthly average of sunspot numbers generated with the uniformly distributed random noise, ε . The solid line represents R_i generated by Eq. (1). The dotted and the dashed lines correspond to R_B generated by Eq. (2) with $\alpha = 0.1$ and $\alpha = 1.0$, respectively. Comparing the resulting daily data sets in Figs. 1 and 2, as concluded in Chang (2008), the exponential noise seems to reproduce the observational features reasonably well.

In Fig. 3, as an example, we show the observed daily number of sunspots (R_B) and R_i for a whole period assuming the exponential random noise with $\alpha = 0.1$. We attempt to apply the linear least-squares method to the relationship between daily R_B and daily R_i to calculate the correction

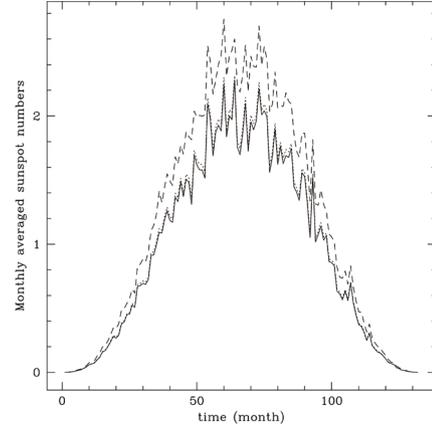


Fig. 1. Monthly average of modeled sunspot numbers generated with the exponentially distributed random noise, ε , as a function of time in a month. The solid line represents R_i generated by Eq. (1). The dotted and the dashed lines correspond to R_B generated by Eq. (2) with $\alpha = 0.1$ and $\alpha = 1.0$, respectively.

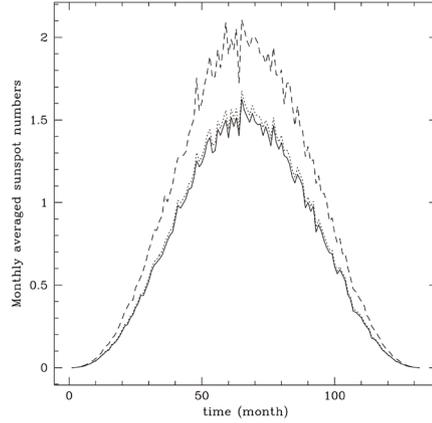


Fig. 2. Similar plots as Fig. 1, but with the uniformly distributed random noise, ε . The solid line represents R_i generated by Eq. (1). The dotted and the dashed lines correspond to R_B generated by Eq. (2) with $\alpha = 0.1$ and $\alpha = 1.0$, respectively.

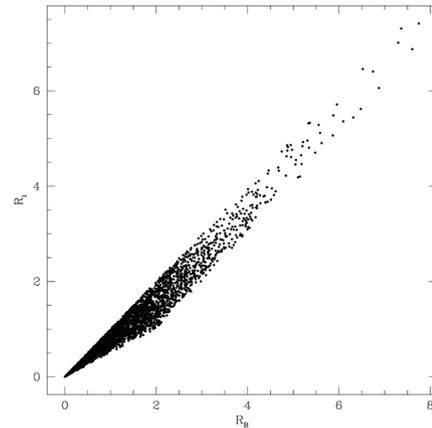


Fig. 3. As an example, relationship between the daily observed number of sunspots (R_B) and R_i for a whole period assuming the exponential random noise with $\alpha = 0.1$.

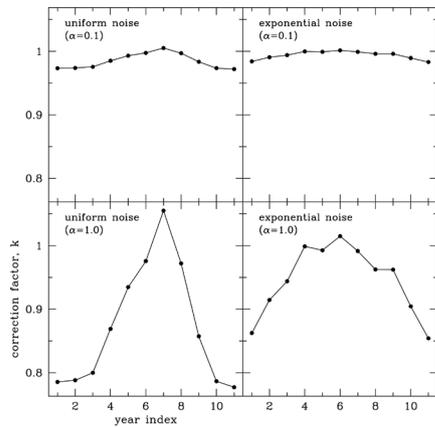


Fig. 4. Resulting yearly k_b values as a function of year. Four cases of two random noise distributions with two different signal-to-noise ratios are indicated at the upper left corner in each panel. Note that the scale in every panel is all the same. The typical uncertainties in determining the slope is the order of 0.01 and 0.001 for $\alpha = 1.0$ and $\alpha = 0.1$, respectively. These values are more or less same for the exponential distribution and for the uniform distribution.

factor k_b in each year, after chopping the whole data set into 11 yearly data subsets. In Fig. 4, we show the resulting yearly k_b values as a function of years since the beginning of the solar cycle. Note that the scale is the same in every panel. Four cases of two random noise distributions with two different signal-to-noise ratios are indicated at the upper left corner in each panel. Resulting correction factors vary from year to year exactly, as was noticed by Oh & Chang (2012). That is, regardless of the noise distribution, we observe that the correction factor increases as the sunspot numbers increase. The typical uncertainties in determining the slope is the order of 0.01 and 0.001 for $\alpha = 1.0$ and $\alpha = 0.1$, respectively. These values are more or less the same for the exponential distribution and for the uniform distribution. The degree of the dependence is subject to the signal-to-noise ratio. Therefore, we consider that the apparent dependence of the value of correction factor k_b on the phase of the solar cycle is not due to a physical mechanism, but is a statistical property of the data. For instance, one may consider the dependence of the correction factor on the phase of the solar cycle since the numbers of large and small sunspots are different, so that the sensitivity, or the detectability of small sunspots, may be the cause of the dependence. If this is the case, as the solar cycle progresses, the size of sunspots could be considered as modulated, and as such the mechanism of sunspot formation under the solar surface is varying in time. What we show here is that this apparent dependence does not require such a complicated mechanism, but can be explained as a statistical property of the observed data. That is, a low number of sunspots results

in a lower correlation coefficient in the R_B-R_i relationship. Furthermore, this effect turns out to be more serious when the noise level is high. When α is large, the resulting scatter (as shown in Fig. 1) gets broader, and as a result the least-squares fit is dominated by noise.

4. DISCUSSION AND CONCLUSIONS

Monitoring sunspots regularly and consistently is an important and basic step in studying various aspects of solar activity. Observers must calculate their own correction factor k regularly and to keep it stable. In South Korea, sunspot observations have been performed at the Korea Astronomy and Space Science Institute (KASI) since 1987 (Sim et al. 1990, Kim et al. 2003) and at the ButterStar Observatory since 2002 (Oh & Chang 2012). All researchers have reported that revisions taking the yearly-basis k into account lead to results that agree better with R_i , and that yearly k apparently varies with the solar cycle. In this paper, using artificial data sets we have modeled, we attempt to investigate whether the variation in the correction factor is really coupled with the solar cycle. We have found that the apparent dependency is due to the statistical property of the data sets. Regardless of statistical distributions of the random noise, we observe the correction factor increases as the sunspot numbers increase. The degree of dependence is also subject to the signal-to-noise ratio. Therefore, we conclude that the apparent dependence of values of the correction factor k on the phase of the solar cycle is not due to a physical mechanism, but is a statistical property of the data.

ACKNOWLEDGMENTS

Authors are grateful to all the past and present members of the ButterStar Observatory for their hard work and dedication, and also thank the anonymous referees for critical comments and helpful suggestions which greatly improve the original version of the manuscript. HYC was supported by the National Research Foundation of Korea Grant funded by the Korean government (NRF-2011-0008123).

REFERENCES

Berghmans D, van der Linden RAM, Vanlommel P, Warnant R, Zhukov A, et al., Solar activity: nowcasting and

- forecasting at the SIDC, *AnGeo*, 23, 3115-3128 (2005).
<http://dx.doi.org/10.5194/angeo-23-3115-2005>
- Chang H-Y, A new method for North-South asymmetry of sunspot area, *JASS*, 24, 261-268 (2007). <http://dx.doi.org/10.5140/JASS.2007.24.4.261>
- Chang H-Y, Stochastic properties in North-South asymmetry of sunspot area, *NewA*, 13, 195-201 (2008). <http://dx.doi.org/10.1016/j.newast.2007.08.007>
- Clette F, Berghmans D, Vanlommel P, van der Linden RAM, Koeckelenbergh A, et al., From the Wolf number to the International Sunspot Index: 25 years of SIDC, *AdSpR*, 40, 919-928 (2007). <http://dx.doi.org/10.1016/j.asr.2006.12.045>
- Kim B-Y, Chang H-Y, Short periodicities in latitudinal variation of sunspots, *JASS*, 28, 103-108 (2011). <http://dx.doi.org/10.5140/JASS.2011.28.2.103>
- Kim RS, Cho K-S, Park Y-D, Moon Y-J, Kim YH, et al., The relative sunspot numbers from 1987 to 2002, *PKAS*, 180, 25-35 (2003). <http://dx.doi.org/10.5303/PKAS.2003.18.1.025>
- Letfus V, Relative sunspot numbers in the first half of eighteenth century, *SoPh*, 194, 175-184 (2000).
- Lockwood M, Twenty-three cycles of changing open solar magnetic flux, *JGR*, 108, 1128-1142 (2003). <http://dx.doi.org/10.1029/2002JA009431>
- Oh S-J, Chang H-Y, Relative sunspot number observed from 2002 to 2011 at Butterstar observatory, *JASS*, 29, 103-113 (2012). <http://dx.doi.org/10.5140/JASS.2012.29.2.103>
- Petrovay K, Solar cycle prediction, *LRSP*, 7, 6-59 (2010).
- Pulkkinen PJ, Brooke J, Pelt J, Tuominen I, Long-term variation of sunspot latitudes, *A&A*, 341, L43-L46 (1999).
- Sim KJ, Kim KM, Park YD, Yoon HS, Analysis of the observational data of sunspots, *PKAS*, 5, 26-39 (1990).
- Solanki SK, Usoskin IG, Kromer B, Schussler M, Beer J, Unusual activity of the Sun during recent decades compared to the previous 11,000 years, *Natur*, 431, 1084-1087 (2004). <http://dx.doi.org/10.1038/nature02995>
- Ternullo M, The butterfly diagram internal structure, *Ap&SS*, 328, 301-305 (2010). <http://dx.doi.org/10.1007/s10509-010-0270-9>
- Usoskin IG, A history of solar activity over millennia, *LRSP*, 5, 3-88 (2008).
- Vaquero JM, Historical sunspot observations: a review, *AdSpR*, 40, 929-941 (2007). <http://dx.doi.org/10.1016/j.asr.2007.01.087>
- Wang Y-M, The Sun's large-scale magnetic field and its long-term evolution, *SoPh*, 224, 21-35 (2004). <http://dx.doi.org/10.1007/s11207-005-4982-x>