Quality Assessment of Sunspot data using different catalogs

Shreya Bhattacharya[1,2], Maarten Jansen[2], Laure Lefèvre[1], Frédéric Clette[1]

1 Royal Observatory of Belgium, WDC-SILSO
2 Université Libre de Bruxelles, Bruxelles, Belgium

ULB UNIVERSITÉ LIBRE DE BRUXELLES

ORB - KSB

WORLD DATA SYSTEM

Abstract

We present a novel time-dependent error determination on sunspot parameters based on non-parametric statistical techniques in smoothing. The overarching goal being the quality assessment of sunspot parameters from different catalogs. In particular we propose a generalized linear regression model with overdispersed count data as response variables in the estimation of a time varying calibration of different sunspot timeseries with overlapping periods.

Introduction

The Sunspot number series is one of the **longest** and most detailed available series in astrophysics spanning over almost four centuries. The series **involves a lot of observers** who differ from each other in terms of their way of counting sunspots, different telescopes and eyesights. Currently the **World Data Centre SILSO** located at the **Royal Observatory of Belgium (ROB)** produces the **International Sunspot Number** and includes 281 contributing stations (Clette et al, 2007).

Most of these stations count only spots and groups, but **some also maintain catalogs of additional sunspot information such as sunspot areas, positions etc.** One such contributing station is the **Uccle Solar Equatorial Table (USET)** situated at ROB. This station has been maintained steadily from **1941 to the present.**

Past studies by Mathieu et al (2019) concluded USET as one of the most stable stations among the network in terms of numbers of groups and spots but the **inclusion of confidence intervals have not yet been achieved consistently**.

In this study we attempted to determine the quality of observations made by the USET station by proposing a model involving non parametric regression. Our model is based on the fact that the **sunspot parameters near the maximum of a solar cycle follows a different probability mass function, from a different random variable than a sunspot parameter near the minimum.**

The overarching goal of this work is to build a **homogeneous series** (Lefevre et al, 2018) of **sunspot parameters** and not just sunspot numbers. In this context, this method proposes a statistically **robust regression method** for determining the calibration factor between observations from different stations. A similar attempt has been made by Mandal et al,(2020) however their approach does not consider the time dependent variability of the sunspot parameters. Since our **proposed model includes time as an explanatory variable** we present a regression coefficient between stations that is dependent on time and is **not restrictive to Sunspot Number only.**

As **this study focuses** on quality assessment of each day observation by USET, we can now produce a **confidence interval for each day of observation.**

Note that, as **it is a work in progress,** we present only the results from three catalogs viz Catania Observatory, USET and the Debrecen Photoheliographic Data (the sunspot catalogue which is compiled as a continuation of Greenwich Photoheliographic (Willis et al. 2013).

Methodology

We present Sunspot data (SN and group area) as overdispersed count data with time as an explanatory variable. We divided the available data in **27 days** which corresponds to one solar rotation (equatorial) and presented the data as:

$$Y(t) \sim F(\mu(t))$$
$$\mu(t) = g^{-1}\{\beta(0,t)\}$$

Where **Y(t)** represents the **data of time series** at day t. F is the probability function which parametrically depends on the **expected value $\mu(t)$, β(0,t) is the intercept of a local (=t-dependent) polynomial** constructed to smooth all observations within the **bandwidth around its center** and **g is the link function**. We propose **an exponential link function.** The expected value $\mu(t)$ has the variance V in the form:

$$V(\mu) = \mu(1+\alpha\mu)$$

such that:

During Minimum: $\alpha \ll 1 \quad V(\mu) = \mu$ **F ~ Poisson**

During Maximum: $\alpha > 1 \quad V(\mu) = \mu(1+\alpha\mu)$ **F ~ Negative Binomial**

Where α is the measure of **overdispersion** determined by **Cameron-Trivedi** criteria (Cameron and Trivedi 1986)

True Value Determination for each day:

For assessing the quality of each day observations, a true value has to be determined for the day. As mentioned, we took three catalogs for this comparison study viz. Catania Observatory Catalog, USET and Debrecan catalog. For each catalog we determined **μ(t),** for each day such that the **daily true value** can be expressed as:

$$y(t) = \frac{\sum_{i=0}^{n} u_i(t)}{n}$$

Where n = number of catalogs used. In this study n=3.

Confidence Band for daily observations:

As mentioned we are focusing our study on observations from USET station. Therefore the quality for each day observation is given by:

$$\text{Residual} = y(t) \sim \mu_{USET}(t)$$

Future Work

A time dependent regression coefficient **b(t)** of the catalogs with respect to each other can be determined by:

b(t) = S(U(t)) S~ Kernel / local polynomial smoothing

Where :

$$\log(U(t) = \log(\mu_1(t)+1) - \log(\mu_2(t)+1)$$

Results

We present the results **for Sunspot Numbers and Sunspot Areas** for the catalogs mentioned.

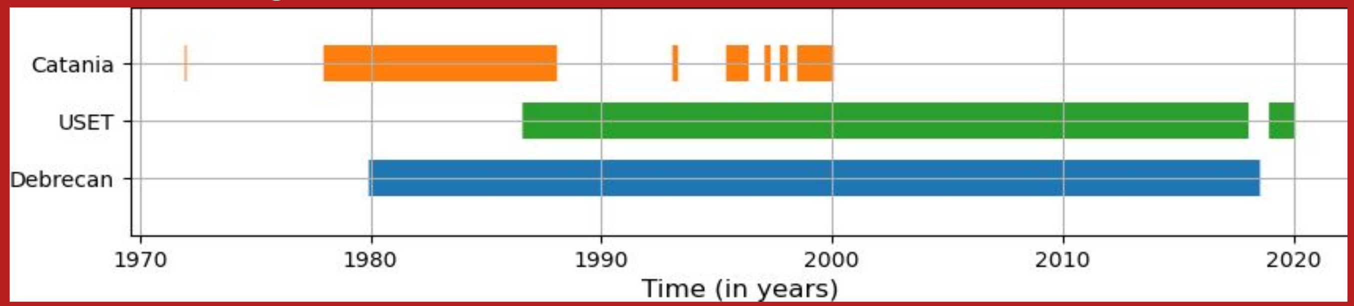

Fig 1: The time spread of the catalogs

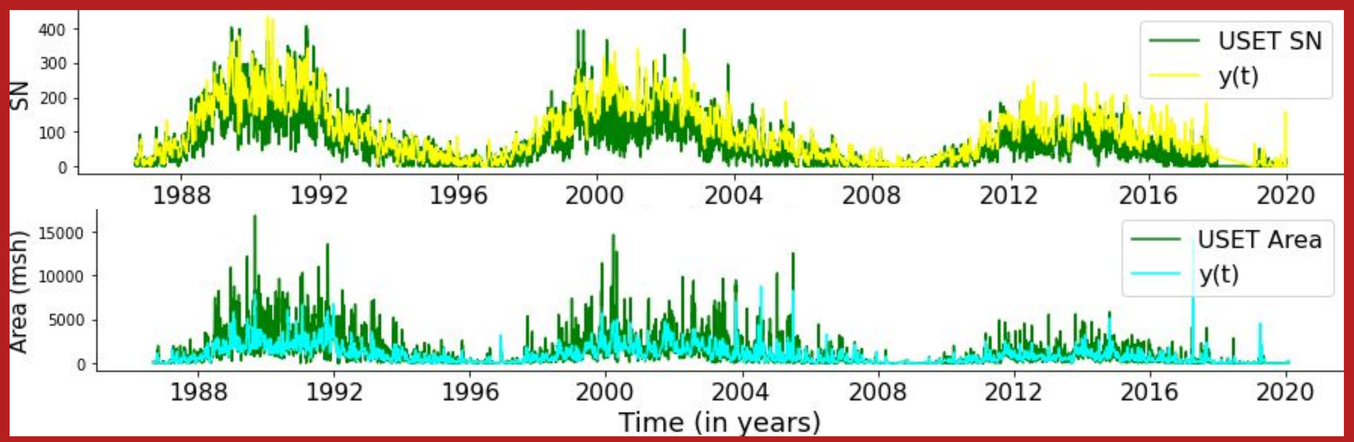

Fig 2: Comparison of the derived daily true value and daily USET observations (SN and Area of Sunspots)

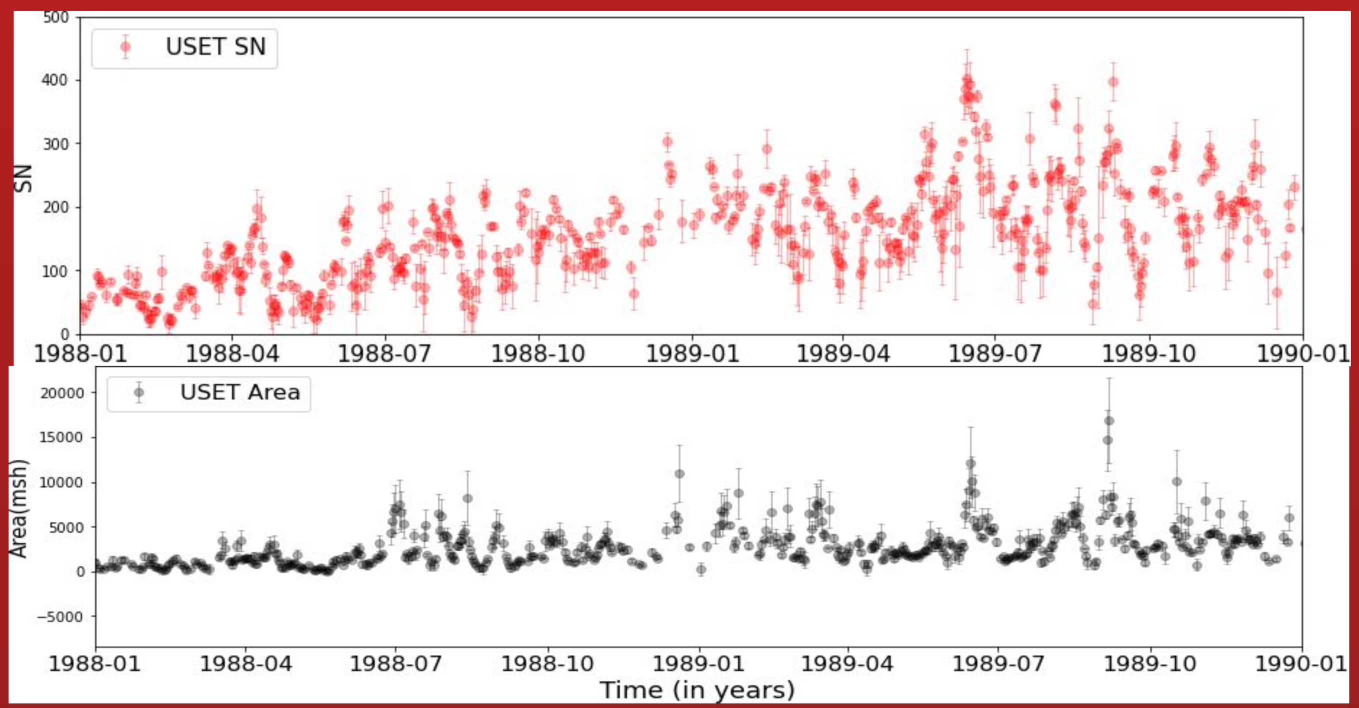

Fig 3: Zoom-in plots representing the confidence levels on daily USET data from 1988-01-01 to 1990-01-01

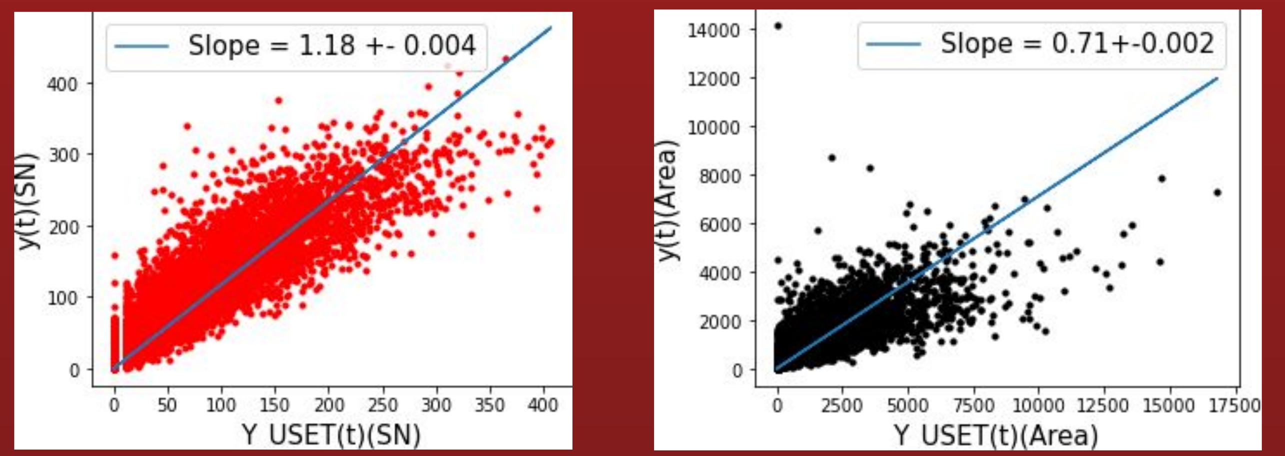

Fig 4: Model fit plot of USET SN and Area observations. The plot shows the fitted values corresponding to original values.

Conclusion

For this study we focused on **only three catalogs** (Fig 1). However, this study can be generalized by including other overlapping catalogs.

Fig 2 confirms the fitted **model follows the trend of the Sunspot Parameters**. We presented the result for only **SN and Area in this study.** However, the study can be expanded to other parameters such as positions , using the same model.

Fig 3 illustrates a sample data having **confidence bands on daily data.**

Fig 4 represents how the original observations of USET corresponds to derived true value. The left panel shows that **USET SN gives an approximate 18% less value than the true daily value**. This result is in fact, in accordance with the result derived by Mathieu et al,2019, where they derived a factor of 0.8 for USET with the network mean. Hence, we confirm **our model does not over/under estimate the counts.**

The right panel shows **USET records almost ~29% more group area** compared to the actual derived value. However, the robustness of the method can be confirmed by including more catalogs.

Acknowledgements

This work was supported by a PhD grant awarded by the Royal Observatory of Belgium to S. Bhattacharya.

References

Clette, F., Berghmans, D., Vanlommel, P., Van der Linden, R. A. M., Koeckelenbergh, A., and Wauters, L., “From the Wolf number to the International Sunspot Index: 25 years of SIDC”, *Advances in Space Research*, vol. 40, no. 7, pp. 919–928, 2007. doi:10.1016/j.asr.2006.12.045

Colin Cameron and Pravin K. Trivedi (1986), "Econometric Models Based on Count Data: Comparisons and Applications of Some Estimators and Tests", *Journal of Applied Econometrics,* January 1986, Vol. 1, pp. 29-54

Lefevre, L., Clette, F., Mathieu, S., Delouille, V., and von Sachs, R., “The Sunspot Number recalibration”, in *Catalyzing Solar Connections*, 2018.

Mandal, S., Krivova, N. A., Solanki, S. K., Sinha, N., and Banerjee, D., “Sunspot area catalog revisited: Daily cross-calibrated areas since 1874”, *Astronomy and Astrophysics*, vol. 640, 2020. doi:10.1051/0004-6361/202037547.

Mathieu, S., von Sachs, R., Ritter, C., Delouille, V., and Lefèvre, L., “Uncertainty Quantification in Sunspot Counts”,*The Astrophysical Journal* vol. 886, no. 1, 2019. doi:10.3847/1538-4357/ab4990.

Willis, D. M., “The Greenwich Photo-heliographic Results (1874 - 1976): Summary of the Observations, Applications, Datasets, Definitions and Errors”, *Solar Physics*, vol. 288, no. 1, pp. 117–139, 2013. doi:10.1007/s11207-013-0311-y.