Contact: daabrand@mtu.edu

Google Scholar

Statistical Methods for Solar Cycle Forecasting: Application to Solar Cycle 25

Daniel A. Brandt[1], Erick F. Vega[1], Brian J. Thelen[1]

[1]Michigan Tech Research Institute, Michigan Technological University

Background

Solar cycle prediction is vital for space mission planning. Solar Cycles are complex and unique (Fig. 1 & 2). There are difficulties using dynamo models to forecast the **solar maximum**; they require unphysical transport coefficients to make accurate predictions. Data-driven methods are thus necessary, but they should be statistically rigorous and incorporate physical understanding. We present one such method, involving the assembling of a large suite of physics-motivated Solar Cycle (SC) statistical features which are down-selected based on information content, and the selected features are fed to Generalized Additive Models that predict the behavior of future Solar Cycles as nonlinear functions of the features.

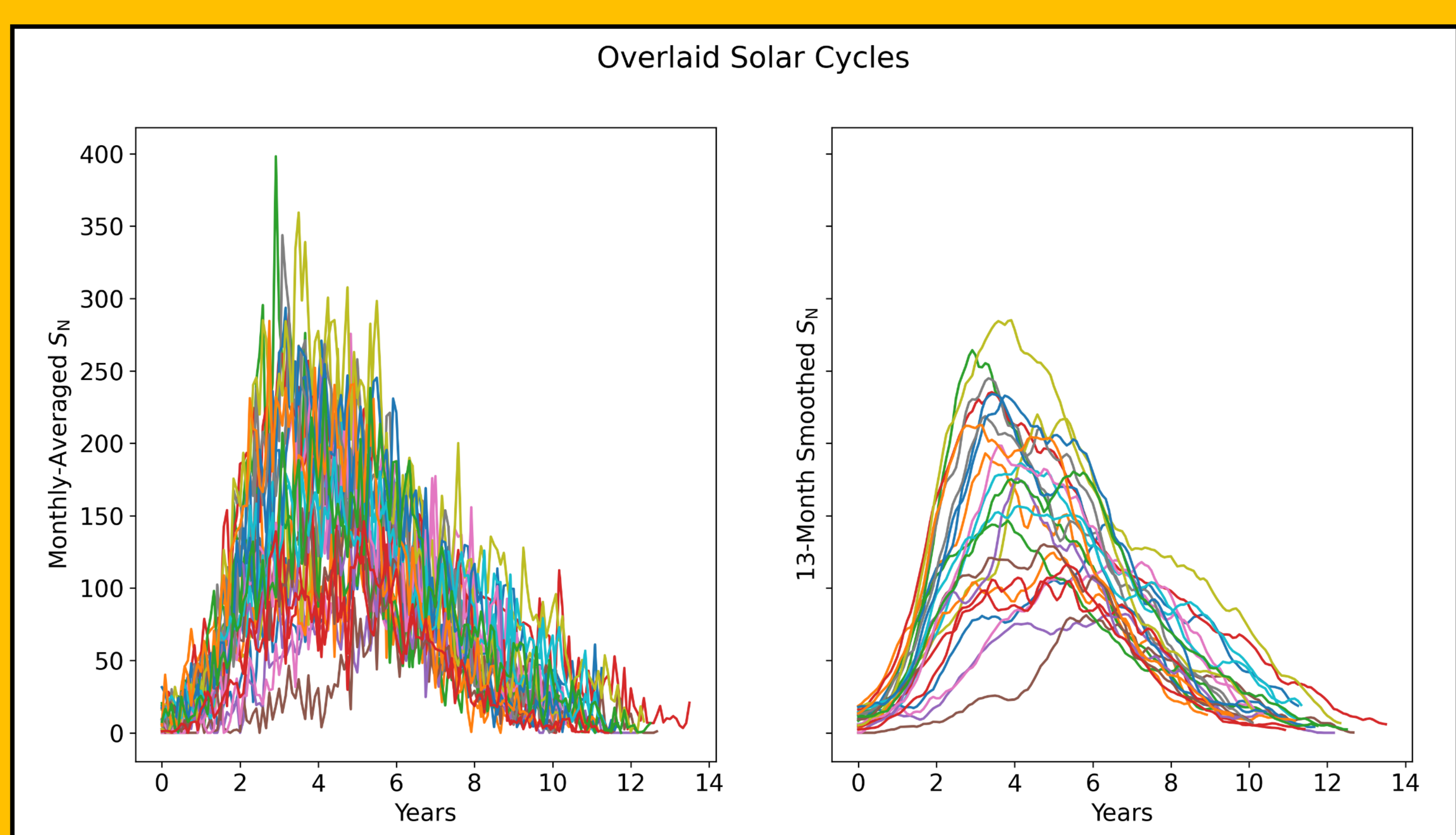

Figure 1: Historical Solar Cycles, plotted on top of each other. Features were extracted from each of these cycles. Each cycle is therefore one sample.

Methodology

Our method proceeds in five steps in two categories:

Feature Selection

1. Assemble physics-motivated features from the Sunspot record.
2. Compute all possible products of the features.
3. Scale the features to the magnitudes of (a) the maximum Sunspot Number in each Solar Cycle ($A_{\max}$) and (b) the time in days from the start of a Solar Cycle to its future maximum ($\tau_{\max}$).
4. Use 'Feature Ordering by Conditional Independence' to downselect the scaled features to the top *N* most relevant for predicting $A_{\max}$ and $\tau_{\max}$.

Model Fitting

5. Use the downselected features as inputs to Generalized Additive Models for predicting $A_{\max}$ and $\tau_{\max}$.

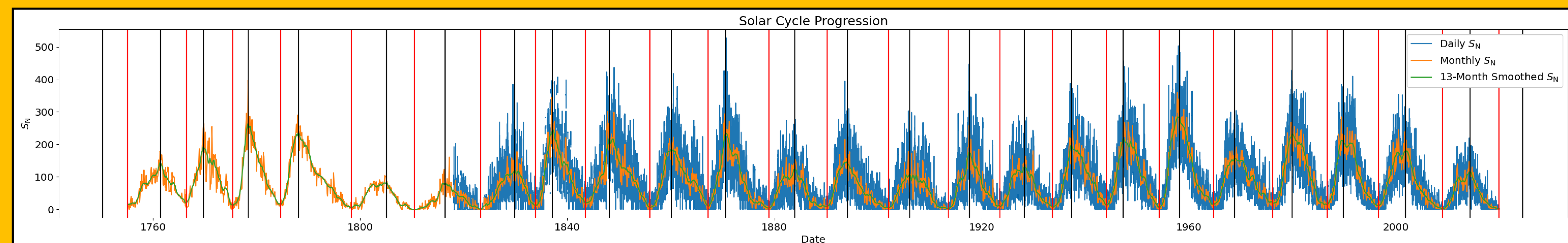

Figure 2: Historical sunspot number (daily, monthly, and 13-monthed smoothed), available from the Royal Observatory of Belgium (https://www.sidc.be/SILSO/datafiles).

Feature Selection

We assemble 9 features *from past cycles*, each with an important physical basis:

- SC max amplitude (*reversal of poloidal field polarity* [1])
- SC min amplitude at SC start (*max strength of poloidal field* [1])
- SC rise time (*fluctuations in meridional circulation* [3])
- SC duration (*meridional flow speed* [2])
- SC rise rate (*fluctuations in meridional circulation* [3])
- SC descent rate (*stochastic fluctuations in poloidal field generation and meridional circulation* [4])
- Ratio between rise and descent rate (*relative strength of stochastic fluctuations in poloidal field*)
- Area-under-the-SC-curve (*poloidal field evolution* [5])
- E-folding time for descending phase *(decay of photospheric field*[6])

We additionally compute all mutual products and use an **affine transformation** to scale the features to the domains of $A_{\max}$ and $\tau_{\max}$.

Feature Ordering by Conditional Independence

- Select the drivers most-correlated to $A_{\max}$ and $\tau_{\max}$ in two steps; downselect all features down to 10, then from 10 down to 7. Use the **coefficient from FOCI [7]** to perform selection:

$$T(Y, \mathbf{Z}|\mathbf{X}) = \int R^2_{Y,\mathbf{Z}|\mathbf{X}} d\nu(t)$$

- FOCI is a **nonlinear generalization of the partial R^2** and has **no tuning parameters**.

Generalized Additive Models

Generalized Additive Models (GAMs) represent the expected value of a response variable as a sum of nonlinear functions of input variables [8]:

$$\mathbb{E}[y|\mathbf{X}] = \beta_0 + \sum_{i=1}^{N} f_i(X_i)$$

We build the f_i from penalized splines, and use our downselected features for $\mathbf{X} = (X_1, X_2, \ldots, X_N)$. The GAM automatically provides confidence intervals for the estimates of $A_{\max}$ and $\tau_{\max}$.

Prediction Results

FOCI+GAMs predict the peak of SC25 to occur between Jan 2024 and Dec 2024 with a maximum of 153±26 with 95% confidence. Current sunspot data suggests the forecast results are reasonable. **Major issues remain regarding the *timing* of the SC maximum.**

FOCI identified the top 3 following features as most relevant for forecasting:

$A_{\max}$: (1) E-folding time, (2) product of Rise Time, Descent Time, and Ratio between the two, and (3) product of Rise Time and Descent Time.

$\tau_{\max}$: (2) Product of SC min amplitude, Rise Time, Descent Rate, and E-folding Time, (2) product of Rise Time, Descent Time, Rise Rate and E-folding Time, and (3) the SC duration.

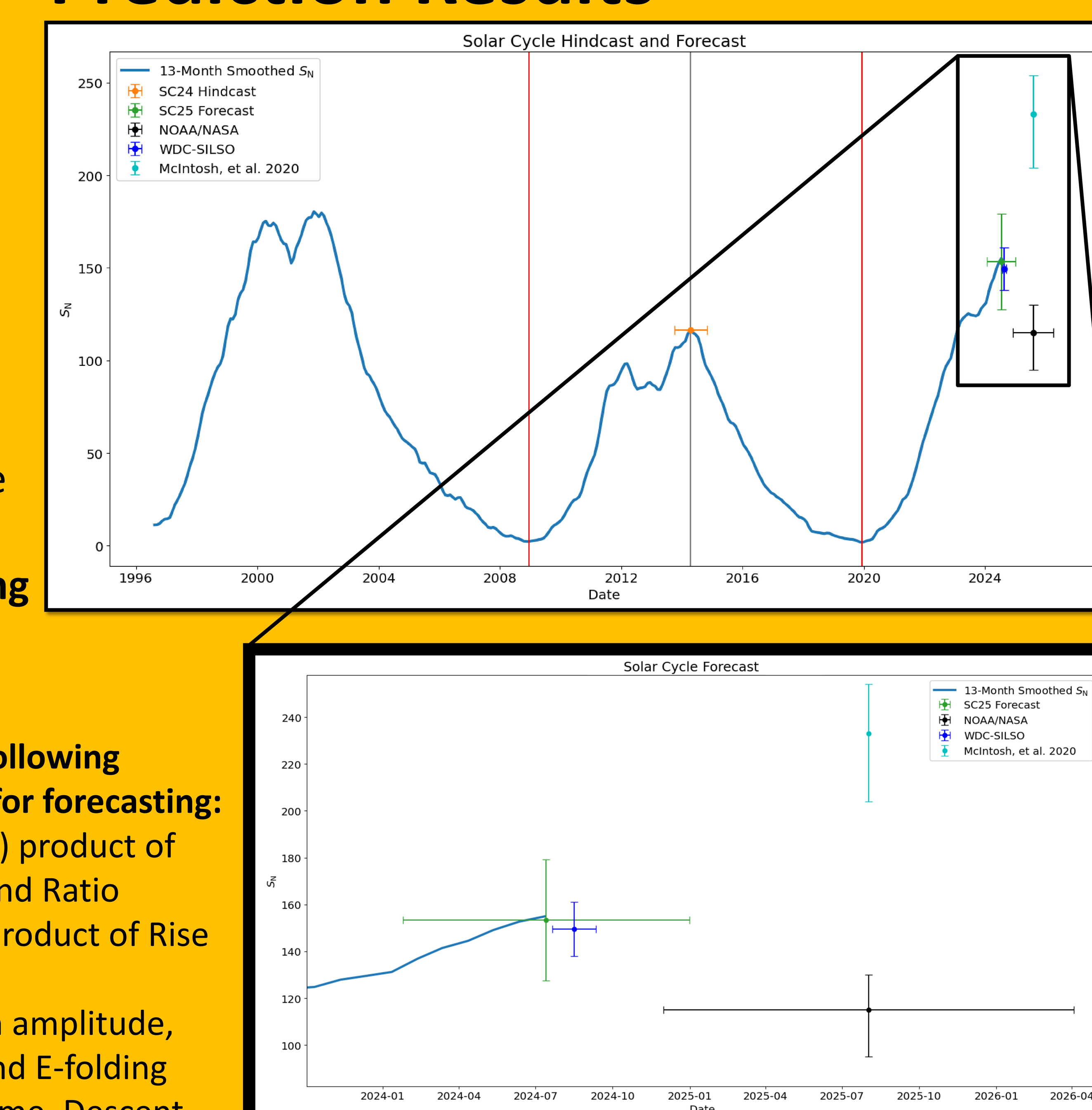

Figure 3: SC25 forecasts generated by FOCI+GAM, NOAA, WDC-SILSO, and [9].

__Implications__: **$A_{\max}$:Photospheric field decay more important than meridional circulation. $\tau_{\max}$: Joint behavior of poloidal field, meridional circulation, and photospheric field most important to forecast timing of maximum; requires better understanding of how rotational shear suppression enables high turbulent diffusion necessary to produce observations ([1], [11]).**

References

[1] Charbonneau, 2020: doi:10.1007/s41116-020-00025-6.
[2] Dikpati and Charbonneau, 1999: doi:10.1086/307269.
[3] Karak and Choudhuri, 2011: doi:10.1111/j.1365-2966.2010.17531.x.
[4] Hazra, et al., 2015: doi:10.1007/s11207-015-0718-8.
[5] Podladchikova, et al., 2008: doi:10.1016/j.astp.2007.08.068.
[6]Wang, et al., 2000: doi:10.1029/1999GL010744.
[7] Azadkia and Chatterjee, 2021: doi:10.1214/21-AOS2073.
[8] Hastie, 1992: ISBN: 9780203738535.
[9] McIntosh, 2020: doi:10.1007/s11207-020-01723-y.
[10] Cameron and Schüssler, 2016: doi:10.1051/0004-6361/201629746.

Acknowledgements

This work was supported in part by the joint NSF and NASA program "Next Generation Software for Data-driven Models of Space Weather with Quantified Uncertainties (SWQU)", under NASA Grant #80NSSC20K1581, along with resources internal to Michigan Tech Research Institute.