Bayesian Analysis of the Solar Cycle Using Multiple Proxy Variables

David C. Stenning[1], David A. van Dyk[2], Yaming Yu[1], Vinay Kashyap[3], C. Alex Young[4]

[1]University of California, Irvine, USA [2]Imperial College London, UK [3]Smithsonian Astrophysical Observatory, USA [4]NASA/GSFC, USA

UCIrvine
Donald Bren School of Information & Computer Sciences

Introduction

Sunspot numbers form a long-duration proxy of solar activity, with records starting in the early seventeenth century. Other proxies of solar activity that have become available more recently show similar patterns and correlations as the sunspot numbers. We extend the Yu *et al.* [1] Bayesian multilevel model of the solar cycle to incorporate data from proxies that have become available more recently, while also taking advantage of the long history of observation of sunspot numbers. Comparing fits of the model using multiple proxies and the sunspot numbers alone reveals significant differences in the inferred cycle properties between the two model fits.

Solar Activity Proxies

We consider three highly correlated proxies of solar activity: *sunspot numbers* (SSNs), *sunspot areas*, and the *10.7cm flux*. A plot of these data is presented in Figure 1. The roughly 11-year cycle of SSNs follows the overall solar cycle. We observe similar patterns in the sunspot areas and the 10.7cm flux, but we also notice differences in the cycle properties implied by the three proxies (e.g., sunspot areas appear to have shorter cycle lengths and less pronounced peaks).

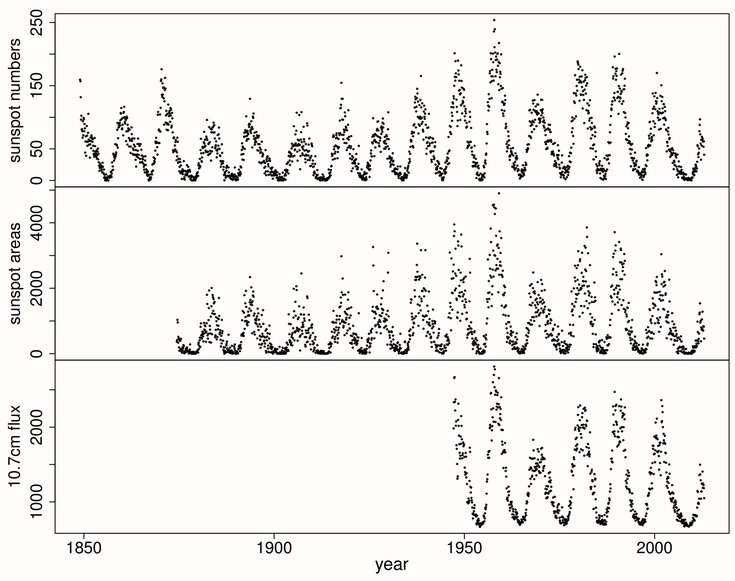

Figure 1 : The observed proxies (monthly averages).

Modeling the Solar Cycle with SSNs

Level One: Modeling the Cycles

Yu *et al.* [1] parameterize the i^{th} solar cycle with a set of cycle-specific parameters: *start time* $t_0^{(i)}$, *time of cycle maximum* $t_{\max}^{(i)}$, *end time* $t_1^{(i)}$, and *amplitude* $c^{(i)}$. The parameterized solar cycle is presented in Figure 2, where $U^{[t]}$ denotes the "average solar activity level" at time t (in months).

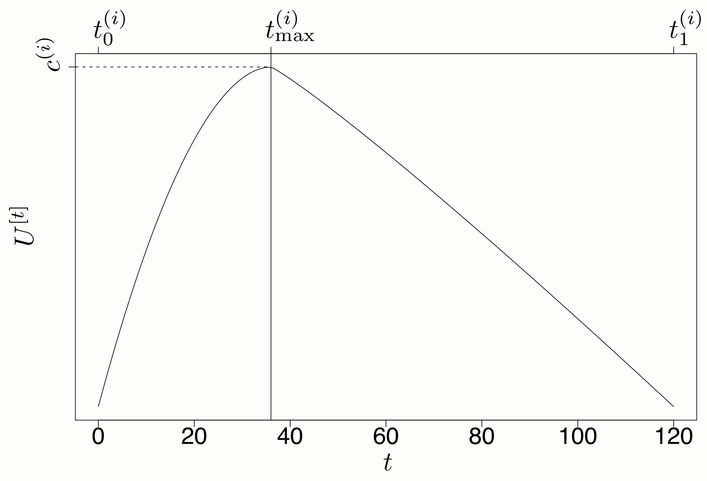

Figure 2 : Parameterized form of a solar cycle.

Level Two: Relationships Between Consecutive Cycles

The evolution of the solar cycle is modeled via a Markov structure on the cycle-specific parameters, see Figure 3. The Markov structure incorporates known features of the solar cycle, such as the Waldmeier effect, and allows for straightforward prediction of ongoing and future cycles.

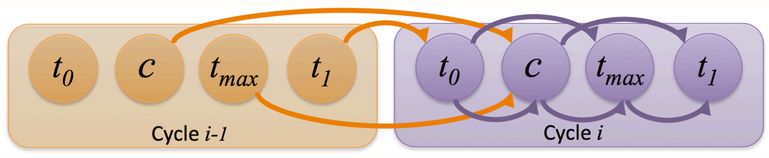

Figure 3 : Markov structure relating the parameters of cycle i to the parameters of cycle $i-1$.

Incorporating Multiple Proxies

A difficulty with combining multiple proxies to model the solar cycle is the varying temporal coverages of the proxies. The SSNs are available as monthly estimates extending back to January 1749, while monthly estimates of sunspot areas and the 10.7cm flux only extend back to May 1874 and February 1947, respectively. There are generally no gaps in the data for an individual proxy once estimates become available, which results in the *monotone missing data pattern* illustrated in Figure 4.

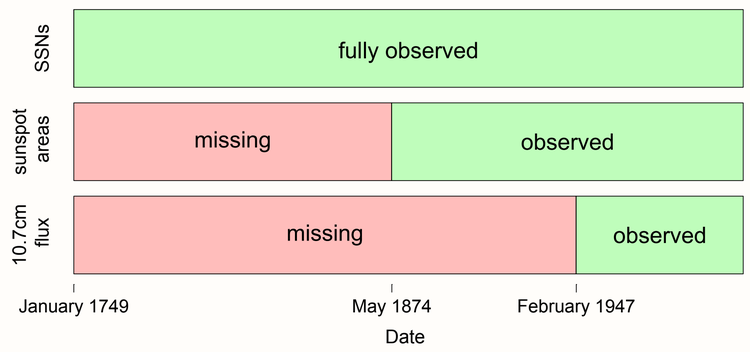

Figure 4 : The monotone missing data pattern. Red bars indicate the time range during which a proxy is missing and green bars indicate the time range during which a proxy is recorded.

Complete-Data Analysis

The proxies exhibit strong linear correlations, see the top row of Figure 5. With no missing data, it is appropriate to use principal component analysis (PCA) to project the multivariate time-series data, Y, onto the one-dimensional manifold defined by the direction of maximum variance. Prior to PCA we use transformations to reduce heteroscedasticity and improve linearity, see the middle row of Figure 5. The univariate time-series data, $G(Y)$, that is then produced via PCA projection represents the overall solar activity level and is highly correlated with the transformed proxy data, see the bottom row of Figure 5. $G(Y)$ is treated as observed data and modeled with the Bayesian multilevel model of the solar cycle that was constructed using the SSNs.

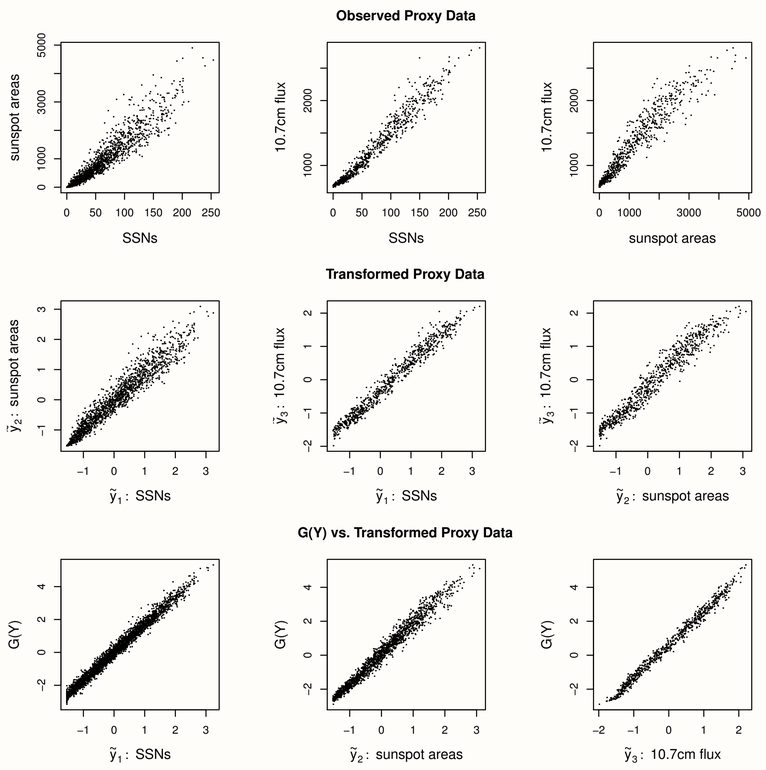

Figure 5 : Scatterplots of observed proxies (Y, top row), transformed proxies ($\tilde{Y}$, middle row), and $\tilde{Y}$ versus $G(Y)$ (bottom row). The transformations are $\tilde{y}_j = \sqrt{y_j^{[t]} + 10}$ for $j = 1, 2$, and $\tilde{y}_3 = \sqrt{y_3^{[t]} - \min_t(y_3^{[t]})}$.

Missing Data

Multiple imputation (MI) [e.g., 2] provides a principled way to use the univariate model to infer the solar cycle using multiple proxies. We specify a local missing data model, where $\tilde{y}_1$ are the SSNs, $\tilde{y}_2$ are the sunspot areas, and $\tilde{y}_3$ is the 10.7cm flux:

$$\tilde{y}_2^{[t]} \mid (\tilde{y}_1^{[t]}, \tilde{y}_2^{[t+1]}) \sim N(\phi_{01} + \phi_{11}\tilde{y}_1^{[t]} + \phi_{21}\tilde{y}_2^{[t+1]},\ \zeta_1) \qquad (1)$$

$$\tilde{y}_3^{[t]} \mid (\tilde{y}_1^{[t]}, \tilde{y}_2^{[t]}, \tilde{y}_3^{[t+1]}) \sim N(\phi_{02} + \phi_{12}\tilde{y}_1^{[t]} + \phi_{22}\tilde{y}_2^{[t]} + \phi_{32}\tilde{y}_3^{[t+1]},\ \zeta_2). \qquad (2)$$

We fit (1) using only the observations for which both $\tilde{y}_1$ and $\tilde{y}_2$ are observed, and likewise for (2). Missing values can then be imputed by drawing from the fitted missing data model.

Results

To allow for comparison we fit the solar cycle model with multiple proxies (multiple-proxy model) and with the SSNs alone (SSN model). With multiple proxies, inference is performed by following the MI combining rules [2] with 5 imputations. The fitted solar cycle for both model fits is displayed in Figure 6. In general, the estimates of $t_{\max}^{(i)}$ under the multiple-proxy model are later than those under the SSN model, although some 95% intervals overlap. We also find that the multiple-proxy model has significantly shorter falling times and total cycle lengths than the SSN model. Prior to the current cycle there is an extended minimum in the fitted solar cycle under the multiple-proxy model, a feature which is absent from the SSN model. Additional results and discussion are available in [3].

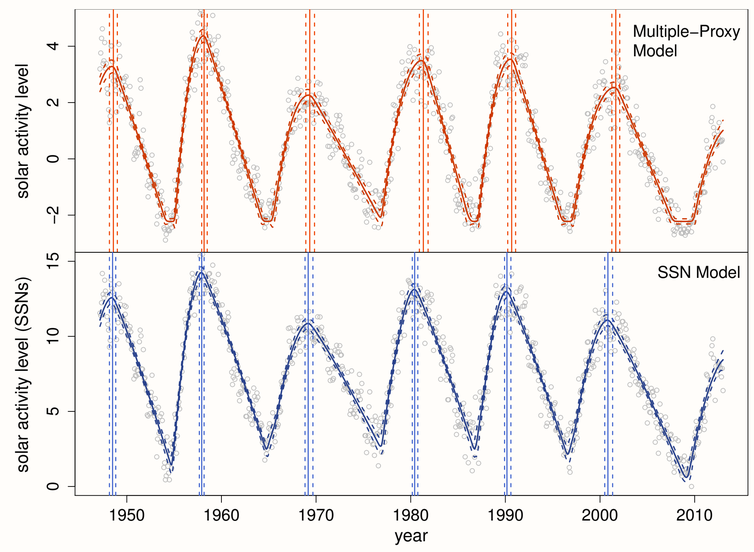

Figure 6 : The fitted solar cycle. *Top:* multi-proxy fit. *Bottom:* SSN fit. The solid (dashed) curves are the fitted solar activity level (95% intervals). The solid (dashed) vertical lines are fitted values for $t_{\max}^{(i)}$ (95% intervals).

Discussion

Our multiple-proxy model of the solar cycle provides the flexibility needed to dynamically describe the complex structure of cycles and their varying shapes, duration, and amplitudes, while capturing the predictable way in which these features evolve over time. Future work will focus on incorporating hemispheric data to capture additional cycle features, such as multiple peaks. As a preliminary step we consider only a single proxy, sunspot areas, since they have the longest history of hemispheric observations. Separate fits of the model using sunspot area data from only the Sun's southern hemisphere (red) and northern hemisphere (blue) are displayed in the top row of Figure 7, and exhibit offsets in several cycle maxima. Using hemispheric sunspot areas results in a smoother overall fit when compared to using only the full-sun sunspot areas, see the middle and bottom row of Figure 7, respectively. Additional proxies, such as solar polar field observations, may further reveal multiple cycle peaks.

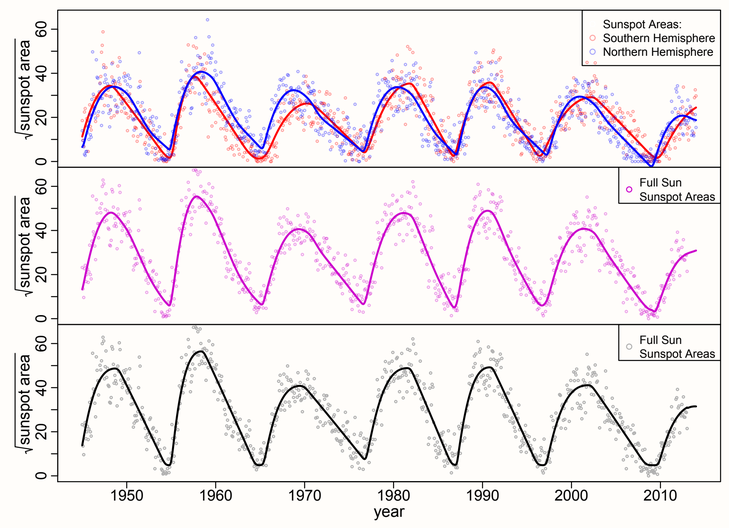

Figure 7 : *Top row*: Hemispheric data and model fits. *Middle row*: Full sun data with (combined) hemispheric model fits. *Bottom row*: Full sun data and model fit.

References

[1] Yu., Y., D.A. van Dyk, V. L. Kashyap, and C.A. Young. (2012). A Bayesian Analysis of the Correlations Among Sunspot Cycles.

[2] Little, R. J. A. and D.B. Rubin. (2002). Statistical Analysis with Missing Data.

[3] Stenning, D.C., D. A. van Dyk, Y. Yu, and V. Kashyap . (2014). A Bayesian Analysis of the Solar Cycle Using Multiple Proxy Variables.