1

A Bayesian Analysis of the Solar Cycle Using Multiple Proxy Variables

David C. Stenning, David A. van Dyk, Yaming Yu, and Vinay Kashyap

Abstract

Solar activity follows a roughly 11-year cyclic pattern that was discovered through proxy variables such as the observed number of sunspots. Correlations between parameters of the consecutive 11-year cycles have also been discovered using sunspot numbers. Other proxies of solar activity, such as sunspot areas and the 10.7cm flux, show similar patterns and correlations as the sunspot numbers. We model solar activity using data from proxies that have become available more recently, while also taking advantage of the long history of observation of sunspot numbers. Yu *et al.* (2012) propose a Bayesian multilevel model of the solar cycle that uses only the sunspot numbers, which we extend to incorporate other proxies. Since proxies have different temporal coverage, we devise a multiple-imputation scheme to account for missing data. We take advantage of strong linear correlations between proxies by using principal component analysis to produce a univariate summary of solar activity at each time point. To examine the effect of incorporating additional proxies we compare fits of the Bayesian multilevel model using multiple proxies and the sunspot numbers alone, finding significant differences in the inferred cycle properties between the two model fits. Specifically, we find that the model fit with multiple-proxies has shorter falling times and shorter overall cycle lengths than the model fit with the sunspot numbers alone.

1.1 Introduction

Highly energetic solar eruptions involving bursts of radiation and discharges of plasma can eject charged particles into space and damage technological infrastructure (*e.g.* radio communications, electric power transmission, and the performance of low-Earth orbit satellites). Such "space weather" events are

common during periods of high *solar activity*—a loose term that is defined only by observable proxy variables. Variations in the level of solar activity follow a roughly 11-year cyclic pattern, which is known as the *solar cycle*. Since energetic space weather events are more common near the *solar maximum*—the peak in solar activity during the 11-year cycle—there is considerable interest in predicting the timing and amplitude of future solar maxima, which has practical value in the planning of space missions. Nevertheless, predicting solar maxima remains a difficult task, with different methods yielding substantially different predictions (see Pesnell, 2012, for an analysis of the various predictions made for the current solar cycle).

The solar cycle was first discovered by observing 11-year cyclic patterns in the average number of *sunspots* visible on the solar disk as viewed from Earth (Wolf, 1852). Sunspots are dark patches on the face of Sun (when viewed in optical light) that occur when intense magnetic fields inhibit convection, temporarily producing areas of reduced surface temperature. Sunspots are therefore linked to the overall magnetic activity of the Sun, and have long been a valuable proxy for solar activity. This value is partly derived from the fact that sunspot numbers (SSNs) comprise the longest uninterrupted set of observations in astronomy, with records starting in the early seventeenth century and available as monthly estimates since 1749. Correlations with solar activity have been established in other proxies such as the 10.7cm flux (i.e. the solar radio flux per unit frequency with a wavelength of 10.7cm), solar flare numbers, sunspot areas, *etc.* (Hudson, 2007). Solar activity can be reconstructed using radiocarbon measurements that are dated using tree-ring data (Bonev *et al.*, 2004; Solanki *et al.*, 2004). Still, the SSNs are the baseline for establishing properties of the solar cycle, and predictions for future cycle maxima are generally based off the SSNs.

Although the oldest SSNs were collected 265 years ago, the data are surprisingly reliable. The sunspot number is $R = k(10g + s)$, where s is the number of individual spots, g is the number of sunspot groups, and k is a factor that corrects for systematic differences between instruments and observatories. Despite advances in technology and the advent of higher resolution images of the Sun, the historical values of R are not expected to have significantly higher uncertainty for two reasons: (1) sunspot size visibility, which affects our ability to see the faintest individual spots, is limited by atmospheric conditions and that limit was reached a long time ago and (2) sunspot groups are always counted as ten individual spots, regardless of the actual number of spots in the group.

Using this data, Waldmeier determined that, within a cycle, the time for SSNs to rise to maximum is less than the time to fall to minimum (Waldmeier, 1935). Other relations, such as the amplitide-period effect—the correlation between the duration of a cycle and the amplitude of the following cycle— were established using SSNs, and can be utilized to predict characteristics of future cycles (e.g., Hathaway *et al.*, 1994, 2002; Watari, 2008). Yu *et al.* (2012) analyzes the SSNs to empirically derive statistically meaningful correlations

between several parameters that they use to describe the solar cycle as part of building a Bayesian multilevel model that accounts for uncertainties in both the average monthly sunspot numbers, and in predicting the characteristics of future cycles. Such correlations help constrain physical models of the *solar dynamo*—the physical mechanism that generates the Sun's magnetic field—that attempt to explain the general solar cycle (Schüssler, 2007).

The statistical model of Yu *et al.* (2012) uses a multilevel structure to capture complex patterns in the solar cycle. The first level of the model parameterizes the solar cycle using using cycle-specific parameters, and describes the distribution of the observed monthly average SSNs around the parameterized cycle. The second level of the model incorporates relationships between the parameters of consecutive cycles, resulting in a hidden Markov model that generates characteristics of cycle $i+1$ given the characteristics of cycle i. While this model was initially fit using observed SSNs, the model can in principle be fit using other proxies that follow the same underlying solar cycle.

A plot of the estimates of monthly average SSNs going back to the mid-nineteenth century is presented in the top panel of Figure 1.1. The cyclic pattern of SSNs that led to the discovery of the solar cycle is clearly visible. The available data extends back to January, 1749, and is maintained and made available by the Solar Influences Data Analysis Center in Belgium (`http://sidc.oma.be`). The bottom two panels of Figure 1.1 present the available data for two additional proxies of solar activity: (1) monthly average total sunspot areas, extending back to May, 1874, also available from the Solar Influences Data Analysis Center and (2) the monthly average 10.7cm flux, which can be obtained from the National Oceanic and Atmospheric Administration's National Geophysical Data Center (`http://www.ngdc.noaa.gov/stp/solar/flux.html`). Although the SSNs, sunspot areas, and the 10.7cm flux have differing temporal coverage, from the period of overlap it is clear that all three of these proxies follow a similar underlying pattern, which is the solar cycle. There are differences, however, in the cycle properties implied by the proxies. For example, the sunspot areas appear to have shorter cycle lengths and less pronounced peaks, especially in the first few observed cycles. The proxies are nonetheless highly correlated and believed to be associated with a common underlying solar cycle. Our goal is to combine information from the multiple proxies into a single omnibus estimate of the underlying cycle.

There are physical explanations as to why these proxies are correlated; the total areas of sunspots will obviously depend on the number of sunspots, and there is evidence that the 10.7cm flux values are influenced by the magnetic fields associated with sunspots (Greenkorn, 2012). Because such proxies have varying temporal coverages and may have varying cadences, it is challenging to combine them to model the underlying solar cycle in a sophisticated statistical analysis. Ideally, we would like to take advantage of the depth of high quality data that has become available in recent years, while incorporating the long record of SSNs.

FIGURE 1.1
The observed proxies. *Top row*: monthly average sunspot numbers. *Middle row*: monthly average sunspot areas. *Bottom row*: monthly average 10.7cm flux. The roughly 11-year cycle of sunspot numbers follows the overall solar cycle, and we observe similar patterns in the sunspot areas and the 10.7cm flux.
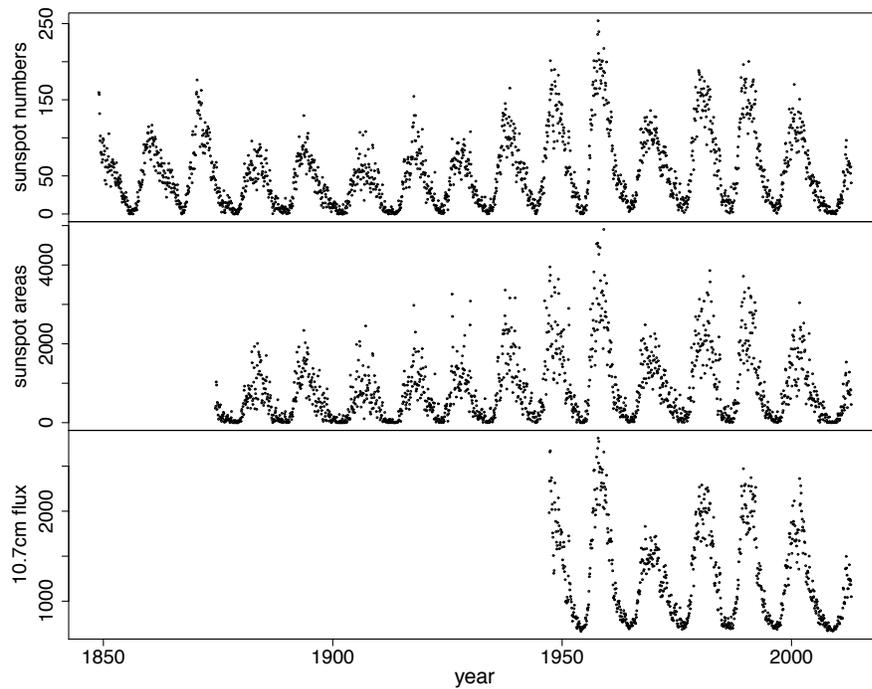
Given the interest in predicting future solar maxima, diverse methodologies have been utilized including those based on i) models of the solar *dynamo*—the physical mechanism that produces the Sun's magnetic field (e.g., Choudhuri, 1992; Charbonneau and Dikpati, 2000; Dikpati and Gilman, 2006; Choudhuri *et al.*, 2007; Charbonneau, 2007), ii) measurements of geomagnetic activity that act as precursors (e.g. Hathaway and Wilson, 2006), and iii) statistical analysis of historical and current data (e.g., Hathaway *et al.*, 1994; Benestad, 2005; Gil-Alana, 2009; Yu *et al.*, 2012), among others. An overview of the various predictions made for the current solar cycle, cycle 24, is given in Pesnell (2012). Within the solar physics community there has been debate over the amplitude of cycle 24, with different physical models producing a large range

of predictions (e.g., Dikpati *et al.*, 2006; Choudhuri *et al.*, 2007). For the previous cycle, cycle 23, Kane (2001) notes that among twenty predictions of the smoothed maximum sunspot number made by different researchers, only eight were within an acceptable range of the later observed value. The Bayesian method of Yu *et al.* (2012) showed that there is considerable uncertainty in predicting an upcoming solar maximum using data up to the current solar minimum.

This chapter is divided into five sections. In Section 2 we review the Yu *et al.* (2012) Bayesian multilevel model for the solar cycle. In Section 3 we develop a systematic strategy for combining multiple proxies, and in particular describe how the pattern of missing data can be exploited to obtain coherent statistical inference. We present the results of fitting the solar cycle model to data that combines multiple proxies in Section 4. Finally, in Section 5 we summarize our results and discuss directions of future research.

## 1.2   Modeling the Solar Cycle with Sunspot Numbers

Yu *et al.* (2012) propose a Bayesian multilevel model for fitting the solar cycle using monthly average SSNs as a proxy for the solar activity level. The solar cycle is parameterized with a set of cycle-specific parameters that together describe the total length, rising time, and amplitude for a given cycle. In the first level of the multilevel model, the observed SSNs are related to the parameterized solar cycle. The second level of the model incorporates a Markov structure that links parameters of consecutive cycles and encapsulates known features of the sunspot cycle in a series of sequential relations. These two stages are combined into a coherent statistical model, which is fit using Markov chain Monte Carlo methods. This structure allows for straightforward prediction of the characteristics of current cycles, even with data only extending to the beginning of the cycle. This is an important facet of the predictive capability of the Bayesian multilevel model.

1.2.1   Level One: Modeling the Cycles

Figure 1.2 illustrates our parameterized model for a single solar cycle. In it, for cycle i, suppose $t_0^{(i)}$ is the starting time, $t_{\max}^{(i)}$ is the time of the cycle maximum, $t_1^{(i)}$ is the end time, $c^{(i)}$ is the amplitude, and $U^{[t]}$ is the "average solar activity level" at time t. Here t is recorded in units of months; although the exact number varies, there are roughly $11 \times 12 = 132$ months per cycle. Under this model, the rising phase of the cycle is described as

FIGURE 1.2
Parameterized form of a solar cycle. We illustrate $U^{[t]}$ with $c^{(i)} = 10, t_0^{(i)} = 0, t_{\max}^{(i)} = 36, t_1^{(i)} = 120$, $\alpha_1 = 1.9$ and $\alpha_2 = 1.1$, where $U^{[t]}$ is specified by (1.1) and (1.2).
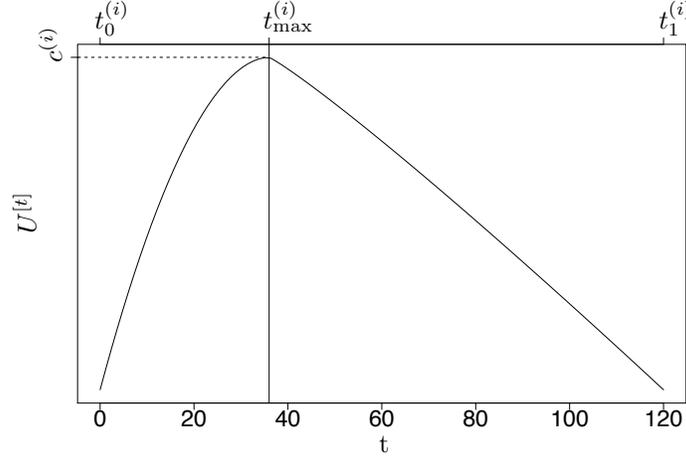
$$U^{[t]} = c^{(i)} \left(1 - \left(\frac{t_{\max}^{(i)} - t}{t_{\max}^{(i)} - t_0^{(i)}} \right)^{\alpha_1} \right) \quad \text{for } t < t_{\max}^{(i)}, \tag{1.1}$$

and the declining phase as

$$U^{[t]} = c^{(i)} \left(1 - \left(\frac{t - t_{\max}^{(i)}}{t_1^{(i)} - t_{\max}^{(i)}} \right)^{\alpha_2} \right) \quad \text{for } t > t_{\max}^{(i)}, \tag{1.2}$$

where $\alpha_1, \alpha_2 > 1$ are shape parameters assumed to be constant for all cycles. Together, (1.1) and (1.2) parameterize the solar cycle, and the curve described by (1.1) and (1.2) is the curve in Figure 1.2. An important feature of this parameterization is that the starting point of the the next cycle, $t_0^{(i+1)}$, is not necessarily identical to the end point of the current cycle, $t_1^{(i)}$. When two cycles overlap (i.e. when $t_1^{(i)} > t_0^{(i+1)}$), the activity level, $U^{[t]}$, is given by the sum of the contributions of the form (1.2) and (1.1) from the respective cycles; when $t_1^{(i)} < t_0^{(i+1)}$, $U^{[t]} = 0$ for $t_1^{(i)} < t < t_0^{(i+1)}$.

The observed SSNs span a total of 25 cycles that are designated cycle 0 through cycle 24. The parameters that are specific to cycle i are the set $\theta^{(i)} = (t_0^{(i)}, t_{\max}^{(i)}, t_1^{(i)}, c^{(i)})$. The collection of cycle-specific parameters is contained in the set $\Theta = (\theta^{(0)}, \ldots, \theta^{(24)})$. Then, the full set of parameters that characterize

the solar cycle is given by (Θ, α), where $\alpha = (\alpha_1, \alpha_2)$ does not vary from cycle to cycle and is therefore not included in Θ. With this, the distribution of the observed SSNs given θ is

$$\sqrt{y}^{[t]} | (\Theta, \alpha, \beta, \sigma^2) \overset{\text{ind}}{\sim} N(\beta + U^{[t]}, \ \sigma^2), \tag{1.3}$$

where $y^{[t]}$ is the monthly average SSN at time t, and the parameter β may be regarded as a baseline. The SSNs are modeled after a square-root transformation in order to stabilize the variance. The independence assumption in (1.3) is valid since sunspots disappear or rotate over the edge of the solar disk over timescales shorter than the observed monthly average SSNs. The independence assumption is not valid when analyzing daily fluctuations since the same sunspot or group of sunspots is counted every day until it vanishes. The decision to use monthly averages was partly motivated to avoid complex modeling of daily correlations.

1.2.2   Level Two: Relationships Between Consecutive Cycles

The evolution of the solar cycle is modeled via a Markov structure on the cycle-specific parameters $\{\theta^{(i)}, i = 0, \ldots, 24\}$. In particular, we model

$$p(\theta^{(0)}, \ldots, \theta^{(24)} | \eta) = p(\theta^{(0)} | \eta) \prod_{i=1}^{24} p(\theta^{(i)} | \theta^{(i-1)}, \eta), \tag{1.4}$$

where η is a set of hyper-parameters that we describe below. The distribution $p(\theta^{(i)} | \theta^{(i-1)}, \eta)$ is further factored in that we model each of the components in cycle i in their temporal order within the cycle. That is, we first model the cycle's start time, $t_0^{(i)}$, given the parameters of the previous cycle, $\theta^{(i-1)}$, then model its amplitude, $c^{(i)}$, given $t_0^{(i)}$ and $\theta^{(i-1)}$, then model the time at which it reaches maximum, $t_{\max}^{(i)}$, given $t_0^{(i)}$, $c^{(i)}$ and $\theta^{(i-1)}$, and finally model its end time, $t_1^{(i)}$, given $t_0^{(i)}$, $c^{(i)}$, $t_{\max}^{(i)}$ and $\theta^{(i-1)}$.

Beginning with the start time of cycle i, Yu *et al.* (2012) allowed the start time, $t_0^{(i)}$, to be different from, but dependent on, the end time of the previous cycle, $t_1^{(i-1)}$. Given $\theta^{(i-1)}$, $t_0^{(i)}$ is modeled as

$$t_0^{(i)} \mid t_1^{(i-1)} \sim t_1^{(i-1)} + N(0, \tau_0^2), \tag{1.5}$$

where the hyper-parameter τ_0^2 regulates the time difference between $t_0^{(i)}$ and $t_1^{(i-1)}$. The conditional distribution of $t_0^{(i)}$ depends on $\theta^{(i-1)}$ only through $t_1^{(i-1)}$.

To formulate $p(c^{(i)}, t_{\max}^{(i)}, t_1^{(i)} | t_0^{(i)}, \theta^{(i-1)}, \eta)$, Yu *et al.* (2012) conducted an exploratory analysis of the observed relationships among the parameters of consecutive cycles. In particular, they fit the model described by (1.3) to each of the 25 cycles individually and used the observed correlations among the

cycle-specific fitted parameters to specify the parametric form of the distribution of $\theta^{(i)}$ given $\theta^{(i-1)}$. For example, there is a positive correlation between consecutive amplitudes $c^{(i)}$ and $c^{(i-1)}$, and a negative correlation between $c^{(i)}$ and $t_0^{(i)} - t_{\max}^{(i-1)}$. The predictive power of the positive correlation between consecutive amplitudes can therefore be enhanced by combining it with the negative correlation with $t_0^{(i)} - t_{\max}^{(i-1)}$. This means that given $\theta^{(i-1)}$, the amplitude of cycle i, $c^{(i)}$, depends on both $c^{(i-1)}$ and $t_0^{(i)} - t_{\max}^{(i-1)}$. Therefore, the distribution of $c^{(i)}$ given $\theta^{(i-1)}$ and $t_0^{(i)}$ is modeled as

$$c^{(i)} \mid (c^{(i-1)}, t_0^{(i)}, t_{\max}^{(i-1)}) \sim \delta_1 + \gamma_1 \frac{c^{(i-1)}}{t_0^{(i)} - t_{\max}^{(i-1)}} + N(0, \tau_1^2). \qquad (1.6)$$

Additional correlations are observed among the components of $\theta^{(i)}$. For example, Yu *et al.* (2012) observe a negative correlation between the rising time of a cycle, $t_{\max}^{(i)} - t_0^{(i)}$, and the amplitude reached during the same cycle. This negative correlation was first discovered by Waldmeier (1935) and is hence known as the "Waldmeier effect." This effect means that the time at which cycle i reaches a peak, $t_{\max}^{(i)}$, is dependent on the starting time and the amplitude of that cycle. The distribution of $t_{\max}^{(i)}$ given $\theta^{(i-1)}$, $t_0^{(i)}$ and $c^{(i)}$ is thus modeled as

$$t_{\max}^{(i)} \mid (t_0^{(i)}, c^{(i)}) \sim t_0^{(i)} + \delta_2 + \gamma_2 c^{(i)} + N(0, \tau_2^2). \qquad (1.7)$$

Notice that $t_{\max}^{(i)}$ is conditionally independent of $\theta^{(i-1)}$.

Finally, Yu *et al.* (2012) observe and incorporate a correlation between the amplitude, $c^{(i)}$, and the time-to-decline, $t_1^{(i)} - t_{\max}^{(i)}$, of that cycle. This means that $t_1^{(i)}$ depends on $c^{(i)}$ and $t_{\max}^{(i)}$. With this, the distribution of $t_1^{(i)}$ given $t_{\max}^{(i)}$ and $c^{(i)}$ is modeled as

$$t_1^{(i)} \mid (t_{\max}^{(i)}, c^{(i)}) \sim t_{\max}^{(i)} + \delta_3 + \gamma_3 c^{(i)} + N(0, \tau_3^2), \qquad (1.8)$$

and $t_1^{(i)}$ is conditionally independent of $\theta^{(i-1)}$.

The relations described by (1.5) to (1.8) can be encapsulated with the Markov structure illustrated in Figure 1.3. This structure allows for straightforward prediction of the characteristics of current cycles, even with data only extending to the beginning of the cycle. This is an important facet of the predictive capability of the Bayesian multilevel model. Together, (1.5) to (1.8) define the joint distribution of $p(\theta^{(i)}|\theta^{(i-1)}, \eta)$, where $\eta = (\tau_0^2, \gamma_j, \delta_j, \tau_j^2, j = 1, 2, 3)$ are the hyper-parameters. Yu *et al.* (2012) also examined correlations between non-adjacent cycles (*i.e.*, between cycle i and cycle $i\pm2$). However, no evidence was found to suggest more than a lag-one dependence. This has important scientific implications since it suggests that the solar dynamo does not retain memory beyond one cycle. Further evidence for this property has been discovered by examining magnetic proxies of solar activity (Muñoz-Jaramillo

FIGURE 1.3

Markov structure relating the cycle-specific parameters $\theta^{(i-1)}$ and $\theta^{(i)}$.
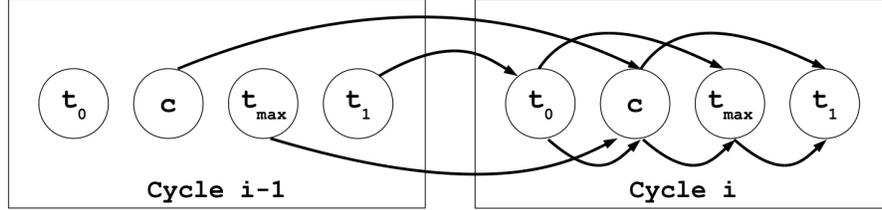
et al., 2013), and by computer simulation of the solar dynamo (Karak and Nandy, 2012).

1.2.3   Prior Distiributions

In (1.3), both β and $\log \sigma$ are given independent uniform prior distributions. To allow a wide range of cycle shapes, a uniform prior distribution on the interval $(1, 3)$ is used for both α_1 and α_2. The cycle-specific parameters for cycle 0, *i.e.* $t_0^{(0)}$, $t_1^{(0)}$, $t_{\max}^{(0)}$, and $c^{(0)}$ are assigned non-informative uniform prior distributions, subject to physical constraints on their ranges. We consider two prior distributions on the hyper-parameters, η, namely $p(\gamma_j, \delta_j, \tau_j^2) \propto \frac{1}{\tau_j}$, $j = 1, 2, 3$, and $p(\gamma_j, \delta_j, \tau_j^2) \propto 1$, $j = 1, 2, 3$. In our numerical analyses, results are not sensitive to this choice and we therefore only report results obtained using $p(\gamma_j, \delta_j, \tau_j^2) \propto \frac{1}{\tau_j}$, $j = 1, 2, 3$.

1.3   Incorporating Multiple Proxies of Solar Activity

An inherent difficulty with combining multiple proxies to model the solar cycle is the varying temporal coverages of the proxies. SSNs, for example, are available as monthly estimates extending back to January, 1749, with no missing data. As technology improved, more proxies began to be observed and, like the SSNs, have been recorded up to the present. For clarity here and in our numerical illustrations, we consider three proxies: monthly average SSNs, monthly average total sunspot areas, and monthly average 10.7cm flux. Estimates of monthly average total sunspot areas extend back to May, 1874. Recordings of the 10.7cm flux began more recently, and estimates of the monthly average are available since February, 1947. It is important to note that, generally, once a proxy comes online (i.e., once a proxy begins to be recorded) it stays online and so there are no gaps in the data for individual proxies, resulting in a

monotone missing data pattern (e.g., Little and Rubin, 2002). This pattern is readily apparent by examining the time-series of the three proxies presented in Figure 1.1, and is described in detail in Section 1.3.2. Before we discuss our strategy for dealing with missing data, we describe how we would handle multiple proxies that were all observed over the same time period.

1.3.1 Complete-Data Analysis

With no missing data, we observe $Y^{[t]} = (y_1^{[t]}, y_2^{[t]}, y_3^{[t]})$ at each t, where y_1 is the monthly average SSN, y_2 is the monthly average sunspot area, and y_3 is the monthly average 10.7cm flux. Since the observed data represent monthly averages, $t = 1, \ldots, 3168$ indexes month. With this scheme, $t = 0$ corresponds to January, 1749, and $t = 3168$ corresponds to December, 2012.

The distribution of $\sqrt{y}^{[t]}$ in (1.3) can be used to model any proxy, perhaps transformed, that follows the same underlying solar cycle. In this way, (1.3) can be generalized to

$$G(Y^{[t]}) \mid (\Theta, \alpha, \beta, \sigma^2) \overset{\text{ind}}{\sim} N(\beta + U^{[t]}, \ \sigma^2), \tag{1.9}$$

where $G(Y^{[t]})$ is a mapping from the multivariate $Y^{[t]}$ to a scalar value, and the underlying parameters are modeled in the same way as in Sections 1.2.2 and 1.2.3. Yu *et al.* (2012) used $G(Y^{[t]}) = \sqrt{y_1}^{[t]}$ to obtain (1.3), but we are interested in finding a $G(Y^{[t]})$ that incorporates information from all available proxies.

The top row of Figure 1.4 displays scatterplots of the observed proxies and illustrates their strong linear correlations. Most of the variability in the data is in one linear dimension. Thus, it is appropriate to employ principal component analysis (PCA) to project the multivariate data onto the one-dimensional manifold defined by the direction of maximum variance.

Before deploying PCA, we transform the proxies to stabilize their variances, which all increase with their mean, see Figure 1.4. For the SSNs and sunspot areas we use the transformation $\sqrt{y_j^{[t]} + 10}$ for $j = 1, 2$. For the 10.7cm flux, we apply the transformation $\sqrt{y_3^{[t]} - \min_t(y_3^{[t]})}$. The constant offsets in the transformations were empirically chosen to improve linearity. We also normalize each transformed proxy by subtracting off its mean and dividing by its standard deviation. Subtracting off the mean is a necessary step in performing PCA, and dividing by the standard deviation controls for the differences in scale between the proxies. We denote the values of the transformed and normalized proxies at time t by $\tilde{Y}^{[t]} = (\tilde{y}_1^{[t]}, \tilde{y}_2^{[t]}, \tilde{y}_3^{[t]})$. Scatterplots of the transformed proxy data, $\tilde{Y}$, are displayed in the second row of Figure 1.4. From these plots we note that the relationships remain linear, the correlations remain strong, but the variances are more stable.

With ω denoting the weights associated with the first principal component

FIGURE 1.4

Transforming the proxies. The top row displays the two-dimensional scatterplots of the observed proxy data, Y. The middle row displays the two-dimensional scatterplots of the transformed proxy data, $\tilde{Y}$. The bottom row displays the computed value of $G(Y) = \omega^T \tilde{Y}$ versus each of the transformed proxies. Notice that $G(Y)$ is highly correlated with each of the transformed proxies.
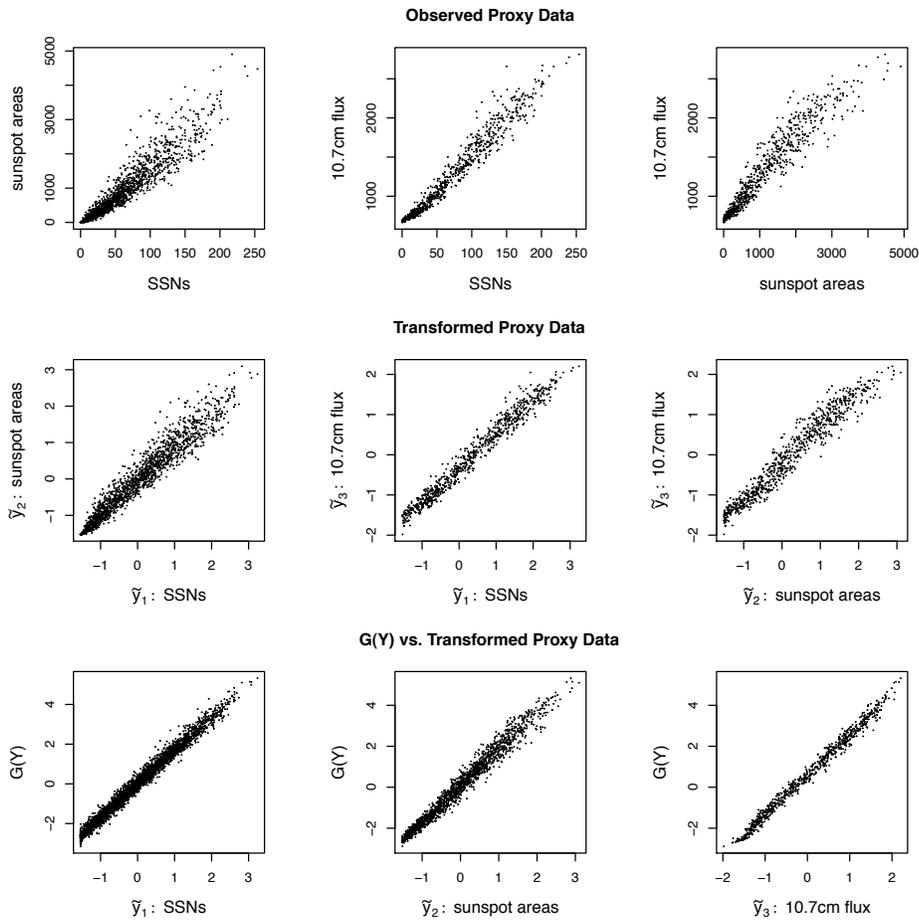
of the $\tilde{Y}^{[t]}$, we let $G(Y^{[t]}) = \omega^T \tilde{Y}^{[t]}$, where $G(Y^{[t]})$ is a scalar value representing the "solar activity level" at time t. The bottom row of Figure 1.4 shows that $G(Y)$ is highly correlated with each of the transformed proxies, which

demonstrates its efficacy as a representation of the overall solar activity level. The first principal component accounts for 98% of the total variability in the proxies, so little information is lost in the PCA-based dimension reduction.

Once the $G(Y^{[t]})$ are obtained they are treated as observed data with distribution given in (1.9). We use a Gibbs sampler to sample from the posterior distribution, incorporating Metropolis-Hastings updates for each conditional draw (see Yu *et al.*, 2012). First, however, we must devise a scheme for handling missing observations of sunspot areas and the 10.7cm flux.

1.3.2   Multiple Imputation Strategy for Missing Data

The monthly average SSNs contain no missing data and are observed for all $\tilde{y}_1^{[t]}$, $t = 1, \ldots, 3168$. Records for the monthly average sunspot areas begin at month $t = 1505$ and contain 1664 observations, and records for the monthly average 10.7cm flux begin at month $t = 2378$ and contain 791 observations. Therefore, the SSNs are observed whenever the sunspot areas are observed, and both the SSNs and sunspot areas are observed whenever the 10.7cm flux is observed. This monotone missing data pattern (see Figure 1.5) allows the development of a straightforward strategy to account for missing data.

Let $\tilde{Y}_{\mathrm{mis}}$ be the missing data, $\tilde{Y}_{\mathrm{mis}} = \{\tilde{y}_2^{[t]}, t = 1, \ldots, 1504; \tilde{y}_3^{[t]}, t = 1, \ldots, 2377\}$, and $\tilde{Y}_{\mathrm{obs}}$ be the observed data. A fully Bayesian strategy for handling the missing data would base inference on $p(\tilde{Y}_{\mathrm{mis}}, \Theta, \alpha, \beta, \sigma, \eta \mid \tilde{Y}_{\mathrm{obs}})$. This could be done by constructing a Gibbs sampler that at each iteration first updates $\tilde{Y}_{\mathrm{mis}} \sim p(\tilde{Y}_{\mathrm{mis}} \mid \tilde{Y}_{\mathrm{obs}}, \Theta, \alpha, \beta, \sigma, \eta)$ and the updates $(\Theta, \alpha, \beta, \sigma, \eta) \sim p(\Theta, \alpha, \beta, \sigma, \eta \mid \tilde{Y}_{\mathrm{mis}}, \tilde{Y}_{\mathrm{obs}})$ using existing computer code. This would, however, require us to specify a model for the multivariate $\tilde{Y}$ rather than simply for the univariate $G(Y)$.

Luckily, multiple imputation (e.g., Little and Rubin, 2002) provides a principled way to use the univariate model and existing computer code to infer the solar cycle using multiple proxies. We first specify a separate simple local missing data model $p(\tilde{Y}_{\mathrm{mis}} | \tilde{Y}_{\mathrm{obs}}, \phi, \zeta)$ that incorporates the Markovian structure inherent in the data . In particular, we model

$$\tilde{y}_2^{[t]} \mid (\tilde{y}_1^{[t]}, \tilde{y}_2^{[t+1]}) \sim N(\phi_{01} + \phi_{11}\tilde{y}_1^{[t]} + \phi_{21}\tilde{y}_2^{[t+1]}, \ \zeta_1) \qquad (1.10)$$

for $t = 1, \ldots, 1504$, and

$$\tilde{y}_3^{[t]} \mid (\tilde{y}_1^{[t]}, \tilde{y}_2^{[t]}, \tilde{y}_3^{[t+1]}) \sim N(\phi_{02} + \phi_{12}\tilde{y}_1^{[t]} + \phi_{22}\tilde{y}_2^{[t]} + \phi_{32}\tilde{y}_3^{[t+1]}, \ \zeta_2) \qquad (1.11)$$

for $t = 1, \ldots, 2377$. Together, (1.10) and (1.11) define the distribution of $p(\tilde{Y}_{\mathrm{mis}} \mid \tilde{Y}_{\mathrm{obs}}, \phi, \zeta)$. We fit (1.10) using only the observations for which both $\tilde{y}_1$ and $\tilde{y}_2$ are observed, and fit (1.11) using only the observations for which all three quantities are observed. With the fitted models in place, $\tilde{Y}_{\mathrm{mis}}$ can be imputed by drawing values from $p(\tilde{Y}_{\mathrm{mis}} \mid \tilde{Y}_{\mathrm{obs}}, \hat{\phi}, \hat{\zeta})$.

Multiple imputation coherently accounts for two sources of uncertainty: the uncertainty that would be present even if all proxies were observed for

FIGURE 1.5

Illustration of the monotone missing data pattern. The solid gray bars indicate the range of observation times t for each of the three solar activity proxies: SSNs, sunspot areas, and 10.7cm flux. White bars indicate the range for which observations are missing. The SSNs are fully observed for months $t = 1, \ldots, 3168$. The sunspot areas are missing for months $t = 1, \ldots, 1504$. The 10.7cm flux is missing for months $t = 1, \ldots, 2377$. SSNs are observed whenever sunspot areas are observed, and both SSNs and sunspot areas are observed whenever the 10.7cm flux is observed. This is a monotone pattern of missing data.

the same time period (*i.e.*, if there was no missing data), and the uncertainty that arises from imputing missing data. With multiple imputation, we obtain M imputations of the missing data, $\tilde{Y}_{\text{mis}}^{(m)} \sim p(\tilde{Y}_{\text{mis}} \mid \tilde{Y}_{\text{obs}}, \hat{\phi}, \hat{\zeta})$ and apply the complete-data analysis described in Section 1.3.1 to each of the imputed data sets. Parameter estimates and uncertainties are based on the multiple imputation combining rules (e.g., Harel and Zhou, 2007; Little and Rubin, 2002). Under this scheme, we obtain M estimates, $\hat{\psi}_m$, of any particular model parameter, ψ, along with their associated variances, V_m. Since we sample from the posterior distribution to fit the Bayesian multilevel model described in Section 1.2, natural candidates for $\hat{\psi}_m$ and V_m are the posterior means and posterior variances of ψ, under each of the M imputed data sets. The multiple-imputation estimate of ψ is $\hat{\psi} = \frac{1}{M} \sum_{m=1}^{M} \hat{\psi}_m$. The estimate of the variance is a combination of the average within-imputation variance, $W = \frac{1}{M} \sum_{m=1}^{M} V_m$, and the between imputation variance, $B = \frac{1}{M-1} \sum_{m=1}^{M} (\hat{\psi}_m - \hat{\psi})^2$, and is

TABLE 1.1
Examining the multiple imputation assumptions. The first column displays the estimates of ζ_1 and ζ_2, along with their estimated standard errors (S.E.), using the *newest third* of the observed data for the sunspot areas and 10.7cm flux, respectively. The second column displays the same, but using the *oldest third* of the observed data for the sunspot areas and 10.7cm flux, respectively. We use an F-test of equality of variances with the null hypothesis that ζ_l for the oldest time epoch is equal to ζ_l for the newest time epoch, $l = 1, 2$. The resulting p-values are given in the third column.

	Newest Third $\hat{\zeta}_1$ (S.E.)	Oldest Third $\hat{\zeta}_2$ (S.E.)	F-test p-value
Imputing Sunspot Areas: $\tilde{y}_2$	3.88 (0.14)	3.17 (0.11)	< 0.01
Imputing 10.7cm Flux: $\tilde{y}_3$	1.37 (0.07)	1.34 (0.07)	0.74

given by $T = W + \frac{M+1}{M}B$. Interval estimates are computed from a reference t-distribution, $(\psi - \hat{\psi})T^{-1/2} \sim t_\nu$, where the degrees of freedom is given by $\nu = (M - 1)\left(1 + \frac{M}{M+1}\frac{W}{B}\right)^2$.

Our multiple imputation procedure relies upon the assumption that ζ_1 and ζ_2 do not vary over time. To test this assumption, we fit the local missing data model for different time periods of data separately and compared the results. We first fit (1.10) using the most recent third of the available $\tilde{y}_2$ data and obtain an estimate and standard error of ζ_1. We then fit (1.10) again using the oldest third of the available $\tilde{y}_2$ data and obtain a second estimate and standard error of ζ_1. We then perform an F-test for the hypothesis that ζ_1^2 for the older time epoch is equal to ζ_1^2 for the newer time epoch. We repeat this procedure for (1.11) using the oldest and most recent third of the $\tilde{y}_3$ data. The results are summarized in Table 1.1. We find that ζ_1 appears to change over time, but that it is larger for the newer data. Since most of $\tilde{y}_2$ is observed, not imputed, we are not particularly worried about overestimating ζ_1 for the older data. In addition, ζ_2 does not appear to vary over time, and this is more important since, unlike $\tilde{y}_2$, the majority of the values of $\tilde{y}_3$ that go into our final analysis are imputed.

1.4   Results

We now discuss the fit of the Bayesian multilevel model for the solar cycle. To allow for comparison we obtain model fits using both $G(Y^{[t]}) = \sqrt{y_1}^{[t]}$

(*i.e.*, the SSN model) and $G(Y^{[t]}) = \omega^T \tilde{Y}^{[t]}$ (*i.e.*, the multiple-proxy model). When $G(Y^{[t]}) = \sqrt{y_1}^{[t]}$, we do not need to perform multiple imputation since there are no missing SSNs. In this case, the fitted values of all quantities are given by their posterior means and credible intervals are given by their 2.5% and 97.5% posterior quantiles. When $G(Y^{[t]}) = \omega^T \tilde{Y}^{[t]}$, we use the multiple imputation strategy described in Section 1.3.2, with $M = 5$ imputations. The estimates for all quantities are computed by setting $\hat{\psi}_m$ and V_m equal to their posterior mean and posterior variances, respectively, and following the multiple imputation combining rules. In particular, the estimate for the solar activity level at time t is computed by setting $\hat{\psi}_m$ and V_m equal to the posterior mean and posterior variance of $U^{[t]} + \beta$. The fitted values of $U^{[t]} + \beta$ are given by the average value of the M within-imputation posterior means and a 95% interval at time t is computed from the reference t-distribution as described in Section 1.3.2.

The fitted values of $U^{[t]} + \beta$ and associated 95% intervals for the multiple-proxy model are plotted in the top panel of Figure 1.6, for the time interval with all three proxies observed (*i.e.,* $t = 2378, \ldots, 3168$). The data presented in this panel are the $\omega^T \tilde{Y}$ values for the given time interval. The bottom panel of Figure 1.6 shows the fitted values of $U^{[t]} + \beta$ and associated 95% intervals under the SSN model for the same time interval; the data are the observed $\sqrt{y_1}^{[t]}$ values. The solid vertical lines in both panels are the fitted values for $t_{\max}^{(i)}$ for cycles $i = 18, \ldots, 23$, and the dashed vertical lines are their 95% intervals. The estimates of $t_{\max}^{(i)}$ under the multiple-proxy model are later than the estimates under the SSN model, although in some cases their 95% intervals overlap. We discuss further comparisons of the timing of the fitted cycles below.

To evaluate the quality of the model fits we plotted the residuals versus time and versus the fitted values, but did not observe any patterns that might call the models into question. The plot of the residuals versus time for the SSN model did not show any significant patterns or evidence of heteroscedasticity that would lead us to question the reliability of the historical sunspot numbers. We also simulate the full time series from the posterior predictive distribution 5000 times. Figure 1.7 displays the 95% posterior predictive intervals from the simulated series along with the $\omega^T \tilde{Y}$ values for the time period when all three proxies are observed. The simulated series are consistent with the observed data.

Since solar physicists are concerned with predicting the timing of solar cycles, we can obtain fitted values and 95% intervals for the rising time of each cycle, $t_{\max}^{(i)} - t_0^{(i)}$, $i = 0, \ldots, 24$, under both the SSN and multiple-proxy models. These are presented in Figure 1.8. The left panel displays the fitted values and 95% intervals for both model fits over time, and the right panel displays a scatterplot of the fitted values under the two models along with their associated 95% intervals. We do not include results for cycle 0 since this initial cycle is incomplete and has relatively large 95% intervals. The 95%

FIGURE 1.6
The fitted solar cycle. The two panels compare the fitted models and data for the multi-proxy (top) and SSN (bottom) models. They include $U^{[t]} + \beta$ (solid curves) and their 95% intervals (dashed curves), along with fitted values for $t_{\max}^{(i)}$ (solid vertical lines) and their 95% intervals (dashed vertical lines). Gray circles represent $\omega^T \tilde{Y}$ (top panel) and $\sqrt{y_1}^{[t]}$ (bottom panel). The time interval displayed covers the period when all three proxies are observed. Using multiple proxies consistently results in later fitted times for the solar maxima.
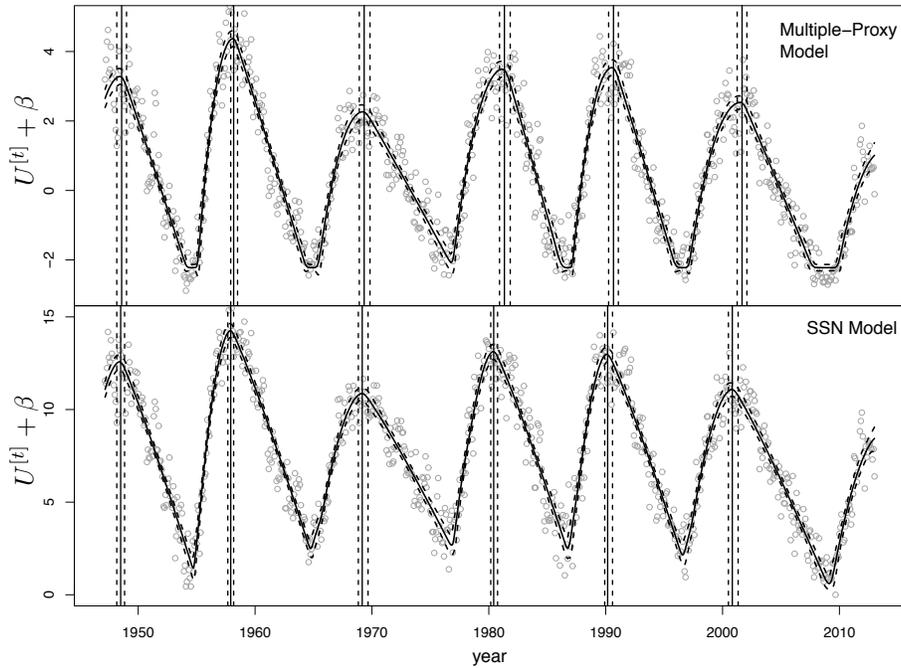

intervals for cycle 24 are also larger than those of other cycles since the cycle is ongoing. That cycle 0 and 24 have fewer observed neighboring cycles than the other cycles also contributes to their larger 95% intervals.

Overall, the rising times do not appear to differ significantly between the two model fits. This is not the case when examining the falling times of each cycle, $t_1^{(i)} - t_{\max}^{(i)}$, which are displayed in Figure 1.9. They are significantly shorter for the model fit with multiple proxies since the 95% intervals rarely intersect the 45° line plotted in the right panel. Taken together, these results suggest that fitting multiple proxies instead of only using the SSNs yields

FIGURE 1.7
Posterior predictive check. The full time series is simulated from the posterior predictive distribution 5000 times, and 95% pointwise posterior predictive intervals are given by the dashed lines. Gray circles represent the observed $G(Y) = \omega^T \tilde{Y}$. The observed data are consistent with the posterior predictive distribution.
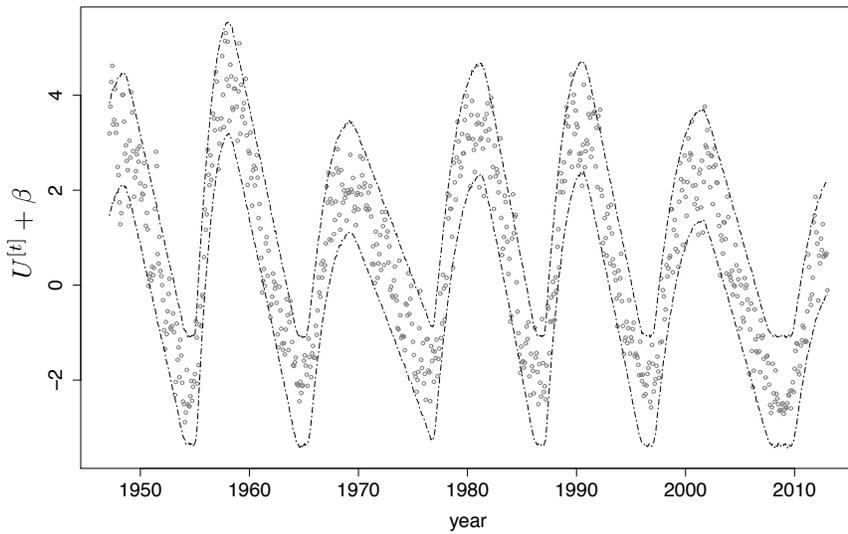

shorter overall cycle lengths, $t_1^{(i)} - t_0^{(i)}$. We display fitted values and 95% intervals for the total cycle lengths in Figure 1.10, and confirm that the model fit with multiple proxies generally has significantly shorter total cycle lengths.

There has been speculation that recent solar cycles represent a period of relatively high activity. Temmer (2010) and Shapoval *et al.* (2013), for example, suggested that the Sun was in a "low-activity" phase from around 1850 to 1915, while the Sun was in a "high-activity" phase from 1915 to the most recent solar minimum. We can evaluate this claim by obtaining an estimate and 95% interval for the mean amplitude during the high-activity phase, cycles 15 though 23, and the low-activity phase, cycles 10 through 14. Let $\bar{c}^{(j:k)}$ represent the mean amplitude from cycles j to k. Then, under the multiple-proxy model, the estimate of $\bar{c}^{(15:23)}$ is 5.01, with 95% interval $(4.86, 5.17)$, and the estimate of $\bar{c}^{(10:14)}$ is 3.86, with 95% interval $(3.70, 4.02)$. Thus, there is evidence that cycles 15 through 23 exhibit higher average solar activity than do cycles 10 through 14. Temmer (2010) and Shapoval *et al.*

FIGURE 1.8
Fitted values and 95% intervals for the cycle rising times under the multiple-proxy and SSN models. There does not appear to be a significant difference in the rising times between the two model fits.
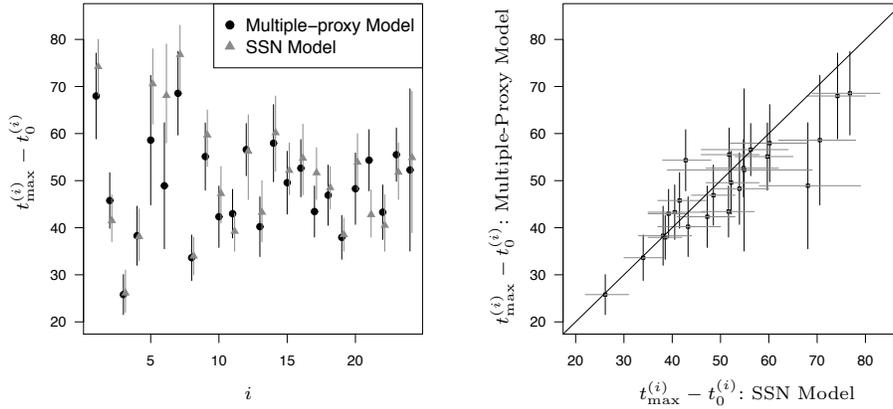

(2013) also hypothesized that cycle 24 represents a shift back to a low-activity phase. Our Bayesian approach allows for straightforward calculation of the probability that the amplitude for cycle 24 will be below $\bar{c}^{(15:23)}$: $Pr(c^{(24)} < \bar{c}^{(15:23)}) = 1$. Futhermore $Pr(c^{(24)} < \bar{c}^{(10:14)}) = 0.88$, which suggests that the amplitude for cycle 24 may be unusually low, even when compared to the low-activity regime.

The Markov structure of our model allows for straightforward prediction of the current cycle even when little data for this cycle is observed. In Figure 1.11, we show the prediction for cycle 24, using data extending up to May 2010 (left panel) and up to December 2012 (right panel). When less data is available from the current cycle, predictions rely more on the Markov structure of the model and are thus more uncertain. As data become available, the predictions are increasingly driven by the current cycle and the uncertainties diminish. The fitted solar cycle is similar in both cases, but the 95% intervals are noticeably narrower with more data. This shows that the cycle-to-cycle relationships we learn in the second stage of the model are consistent with the most recent data and we can make reasonable predictions of cycle characteristics at the start of the cycle, albeit with considerable uncertainty.

FIGURE 1.9
Fitted values and 95% intervals for the cycle falling times under the multiple-proxy and SSN models. Unlike the cycle rising times, the falling times do appear to differ between the two model fits. Specifically, the falling times are significantly shorter under the model fit with multiple proxies.
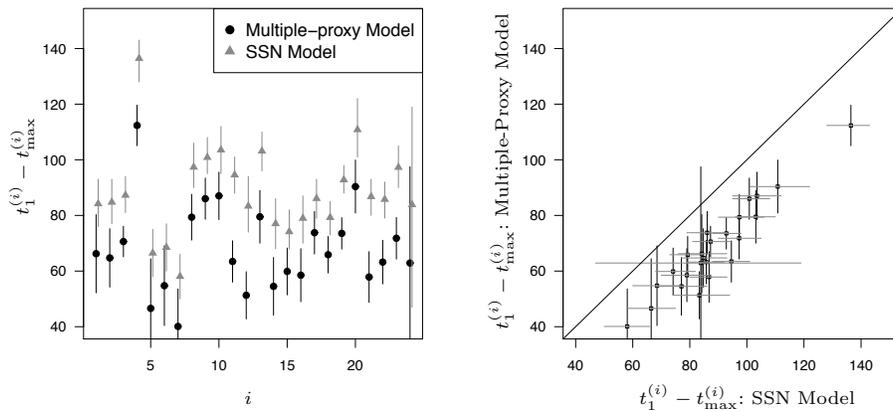

1.5   Summary and Discussion

We have carried out a fully Bayesian analysis of the solar cycle using multiple proxy variables by generalizing the model of Yu *et al.* (2012). After suitably transforming the data to stabilize the variance and increase linearity, we multiply-imputed missing data by specifying a simple local missing data model that incorporates the Markovian structure of the data. The dimensionality of each imputed data set is reduced using PCA to project the multivariate proxy observations onto a one-dimensional subspace along the direction of highest variance. In this way, we obtain a univariate summary of solar activity at each time point, allowing us to utilize the existing univariate model to infer properties of the solar cycle using multiple proxies. This approach is based on the current understanding that there is a single underlying solar cycle, and the several proxies all provide information about the cycle.

It is necessary to use the long history of the SSNs and sunspot area observations in order to learn the patterns among consecutive solar cycles. Multiple imputation is used in order to easily derive estimates based on the posterior distribution of the model parameters given all of the data. If only complete observations are used, meaning only the data with all three proxies observed,

FIGURE 1.10
Fitted values and 95% intervals for the total cycle lengths under the multiple-proxy and SSN models. Following from the results displayed in Figures 1.8 and 1.9, the multiple-proxy model generally has significantly shorter cycle lengths than the SSN model has.
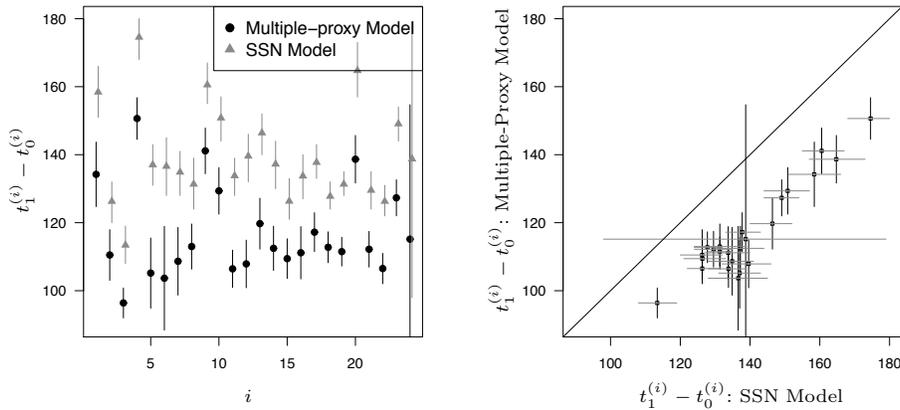

the model is overfit. This problem would be compounded in future analyses that may also include other proxies that do not extend back for more than a couple of cycles. One of our primary aims is to allow additional proxies, even if they are largely missing, so that all available data can be used in a coherent statistical framework.

We compare fits of the Bayesian multilevel model of the solar cycle based on (ii) multiple proxies and (ii) the SSNs alone. We observe significant differences in the inferred cycle properties. In particular, we find that the model fit with multiple proxies has shorter falling times than the model fit with the SSNs. Since we do not find significant differences in the rising times, the shorter falling times from the multiple-proxy model also imply shorter total cycle lengths, which we also observe. Shorter cycle lengths is turn imply longer solar minima. It has been observed elsewhere that the Sun can remain in a prolonged state of minimum activity, and there is evidence that the most recent solar minimum was unusual in its depth and duration (Basu, 2013). During the most recent solar minimum the 10.7cm flux was the lowest ever recorded, and physical characteristics of the solar surface and interior were unusual when compared to previous solar minima (Basu, 2013). It is clear from Figure 1.6 that the most recent minimum of the fitted solar cycle has a longer duration under the multiple-proxy model. In this regard the multiple-

FIGURE 1.11
Prediction for cycle 24, using data up to May 2010 (left) and up to December 2012 (right). The solid curve is the fitted $U^{[t]} + \beta$, and 95% posterior (predictive) intervals are given by the dotted lines. Gray circles represent the observed $G(Y) = \omega^T \tilde{Y}$. As more data for the cycle is obtained, the uncertainty in the predictions is reduced.
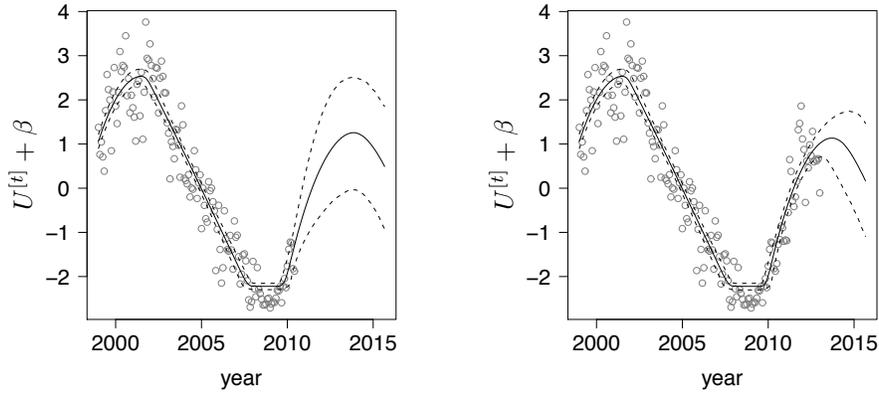
proxy fit captures an important feature in the solar cycle that is missed by the SSN model. The use of additional proxies may further illuminate this effect.

Future work will also consider additional functional forms for the solar cycle. Recent studies have presented evidence that many solar cycles have double maxima (e.g., Georgieva, 2011; Kilcik and Ozguc, 2014), which our current parameterization does not capture. Gnevyshev (1967) suggests that complex physical processes produce double peaks in all cycles, but often the gap between them is too short for the two within-cycle maxima to be distinguished. One possible explanation for double maxima that we can explore is the existence of separate cycles acting on each hemisphere of the Sun. Under this scenario the separate northern-hemisphere and southern-hemisphere cycles are parameterized by (1.1) and (1.2), with total activity being the sum over the two hemispheres and double maxima appearing when the two hemispheres reach peak activity at different times.

Our multiple-proxy Bayesian multilevel 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. The effective combination of multiple imputation and PCA-based dimension reduction makes it

straightforward to incorporate additional proxies, all the while taking advantage of the long history of SSN observations.

Bibliography

Basu, S. (2013). The peculiar solar cycle 24 - where do we stand? *Journal of Physics Conference Series*, **440**(1), 012001.

Benestad, R. E. (2005). A review of the solar cycle length estimates. *Geophysical Research Letters*, **32**, 15714.

Bonev, B. P., Penev, K. M., and Sello, S. (2004). Long-Term Solar Variability and the Solar Cycle in the 21st Century. *The Astrophysical Journal Letters*, **605**, L81–L84.

Charbonneau, P. (2007). Babcock Leighton models of the solar cycle: Questions and issues. *Advances in Space Research*, **39**, 1661–1669.

Charbonneau, P. and Dikpati, M. (2000). Stochastic Fluctuations in a Babcock-Leighton Model of the Solar Cycle. *The Astrophysical Journal*, **543**, 1027–1043.

Choudhuri, A. R. (1992). Stochastic fluctuations of the solar dynamo. *Astronomy and Astrophysics*, **253**, 277–285.

Choudhuri, A. R., Chatterjee, P., and Jiang, J. (2007). Predicting Solar Cycle 24 With a Solar Dynamo Model. *Physical Review Letters*, **98**(13), 131103.

Dikpati, M. and Gilman, P. A. (2006). Simulating and Predicting Solar Cycles Using a Flux-Transport Dynamo. *The Astrophysical Journal*, **649**, 498–514.

Dikpati, M., de Toma, G., and Gilman, P. A. (2006). Predicting the strength of solar cycle 24 using a flux-transport dynamo-based tool. *Geophysical Research Letters*, **33**, 5102.

Georgieva, K. (2011). Why the Sunspot Cycle Is Double Peaked. *ISRN Astronomy and Astrophysics*, **2011**.

Gil-Alana, L. A. (2009). Time Series Modeling of Sunspot Numbers Using Long-Range Cyclical Dependence. *Solar Physics*, **257**, 371–381.

Gnevyshev, M. (1967). On the 11-years cycle of solar activity. *Solar Physics*, **1**(1), 107–120.

Greenkorn, R. A. (2012). A Comparison of the 10.7-cm Radio Flux Values and the International Sunspot Numbers for Solar Activity Cycles 19, 20, and 21. *Solar Physics*, **280**, 205–221.

Harel, O. and Zhou, X.-H. (2007). Multiple imputation: review of theory, implementation and software. *Statistics in Medicine*, **26**(16), 3057–3077.

Hathaway, D. H. and Wilson, R. M. (2006). Geomagnetic activity indicates large amplitude for sunspot cycle 24. *Geophysical Research Letters*, **33**, 18101.

Hathaway, D. H., Wilson, R. M., and Reichmann, E. J. (1994). The shape of the sunspot cycle. *Solar Physics*, **151**, 177–190.

Hathaway, D. H., Wilson, R. M., and Reichmann, E. J. (2002). Group Sunspot Numbers: Sunspot Cycle Characteristics. *Solar Physics*, **211**, 357–370.

Hudson, H. S. (2007). The Unpredictability of the Most Energetic Solar Events. *The Astrophysical Journal Letters*, **663**, L45–L48.

Kane, R. P. (2001). Did Predictions of the Maximum Sunspot Number for Solar Cycle 23 Come True? *Solar Physics*, **202**, 395–406.

Karak, B. B. and Nandy, D. (2012). Turbulent Pumping of Magnetic Flux Reduces Solar Cycle Memory and thus Impacts Predictability of the Sun's Activity. *The Astrophysical Journal Letters*, **761**, L13.

Kilcik, A. and Ozguc, A. (2014). One possible reason for double-peaked maxima in solar cycles: Is a second maximum of solar cycle 24 expected? *Solar Physics*, **289**(4), 1379–1386.

Little, R. J. A. and Rubin, D. B. (2002). *Statistical Analysis with Missing Data*. Wiley Series in Probability and Statistics. Wiley, New York, 2nd edition.

Muñoz-Jaramillo, A., Dasi-Espuig, M., Balmaceda, L. A., and DeLuca, E. E. (2013). Solar Cycle Propagation, Memory, and Prediction: Insights from a Century of Magnetic Proxies. *The Astrophysical Journal Letters*, **767**, L25.

Pesnell, W. D. (2012). Solar cycle predictions (invited review). *Solar Physics*, **281**(1), 507–532.

Schüssler, M. (2007). Are solar cycles predictable? *Astronomische Nachrichten*, **328**, 1087.

Shapoval, A., Le Mouël, J. L., Courtillot, V., and Shnirman, M. (2013). Two Regimes in the Regularity of Sunspot Number. *The Astrophysical Journal*, **779**, 108.

Solanki, S. K., Usoskin, I. G., Kromer, B., Schüssler, M., and Beer, J. (2004). Unusual activity of the Sun during recent decades compared to the previous 11,000 years. *Nature*, **431**, 1084–1087.

Temmer, M. (2010). Statistical Properties of Flares and Sunspots over the Solar Cycle. In S. R. Cranmer, J. T. Hoeksema, and J. L. Kohl, editors, *SOHO-23: Understanding a Peculiar Solar Minimum*, volume 428 of *Astronomical Society of the Pacific Conference Series*, page 161.

Waldmeier, M. (1935). *Astron. Mitt. Eidgen. Sternw. Zürich*, **133**(105).

Watari, S. (2008). Forecasting solar cycle 24 using the relationship between cycle length and maximum sunspot number. *Space Weather*, **6**(12).

Wolf, R. (1852). *Viertel. Natur. Ges. Bern*, **254**(179).

Yu, Y., van Dyk, D. A., Kashyap, V. L., and Young, C. A. (2012). A Bayesian Analysis of the Correlations Among Sunspot Cycles. *Solar Physics*, **281**, 847–862.