Highlights of Spanish Astrophysics XI, Proceedings of the XV Scientific Meeting of the Spanish Astronomical Society held on September 4–9, 2022, in La Laguna, Spain. M. Manteiga, L. Bellot, P. Benavidez, A. de Lorenzo-Cáceres, M. A. Fuente, M. J. Martínez, M. Vázquez- Acosta, C. Dafonte (eds.), 2023

A probability distribution for the amplitude of Solar Cycle 25.

Arregui, I.[1,2]

[1] Instituto de Astrofísica de Canarias, Vía Láctea S/N, E-38205 La Laguna, Spain
[2] Departamento de Astrofísica, Universidad de La Laguna, E-38206 La Laguna, Spain

Abstract

We suggest a method to calculate the probability for the maximum amplitude of Solar Cycle 25 using Bayesian inference. We illustrate this approach with the predictions made by one particular phenomenological model that relates the time interval between termination events of preceding cycles with the amplitude of the next cycle. Our results show well-constrained posterior-predictive distributions for the maximum sunspot-number. The impact of uncertainty on sunspot-number and time interval between terminators is quantified. A comparison between past maximum sunspot-number values and posterior-predictive distributions computed using the method enables us to quantify the quality of the inference and of the prediction.

1 Introduction

Solar cycle prediction has been a matter of interest for decades and has led to a vast literature, see for example the recent reviews by [3, 6, 5]. Among the different techniques employed, precursor methods aim at predicting the amplitude of a given cycle based on a measure of solar activity/magnetism in a preceding cycle at a given moment of time.

One example is the empirical relationship between the time interval between termination events and the amplitude of the upcoming solar cycle, recently suggested by [4]. Termination events delimit epochs of toroidal-magnetic-activity-band interaction and mark the limit between 11-year sunspot cycles and the end of 22-year magnetic activity cycles. According to the proposed relationship, widely separated terminators would correspond to low-amplitude sunspot cycles. Conversely, more narrowly separated terminators would lead to large amplitude sunspot-cycles.

A drawback of the statistical method employed by [4] and similar studies is that they do not permit one to make probability statements, nor do they offer straightforward ways to propagate the uncertainty from the observations to the quantities of interest. We propose a method for computing the probability distribution of the maximum amplitude of Solar Cycle 25 using Bayesian inference and illustrate the method using the phenomenological model and data by [4].

2 Method

Given a model $\mathcal{M}$, with parameter vector $\boldsymbol{\theta}$, proposed to explain observed data $\mathcal{D}$, the posterior distribution of the parameters is given by the product of the likelihood function $p(\mathcal{D}|\boldsymbol{\theta},\mathcal{M})$ and the prior distribution $p(\boldsymbol{\theta}|\mathcal{M})$

$$p(\boldsymbol{\theta}|\mathcal{D},\mathcal{M}) = \frac{p(\mathcal{D}|\boldsymbol{\theta},\mathcal{M})\, p(\boldsymbol{\theta}|\mathcal{M})}{p(\mathcal{D}|\mathcal{M})}. \tag{1}$$

The quantity in the denominator is the evidence, or prior predictive distribution

$$p(\mathcal{D}|\mathcal{M}) = \int_{\boldsymbol{\theta}} p(\mathcal{D}|\boldsymbol{\theta},\mathcal{M})\, p(\boldsymbol{\theta}|\mathcal{M})\, d\boldsymbol{\theta}. \tag{2}$$

The evidence is a measure of the quality of the model. It quantifies how well the data $\mathcal{D}$ are predicted by the model $\mathcal{M}$.

Once the inference of the model parameters is performed, a distribution over possible unobserved future data $\widetilde{\mathcal{D}}$, conditional on the observed data and the inferred model, is given by the posterior-predictive distribution

$$p(\widetilde{\mathcal{D}}|\mathcal{D},\mathcal{M}) = \int_{\boldsymbol{\theta}} p(\widetilde{\mathcal{D}}|\boldsymbol{\theta},\mathcal{M})\, p(\boldsymbol{\theta}|\mathcal{D},\mathcal{M})\, d\boldsymbol{\theta}. \tag{3}$$

The first factor in the integrand is the likelihood of the new unobserved data as a function of the parameter vector. The second factor is the posterior inferred from the old observed data.

3 Analysis and results

Following [4], we assume a linear relationship between the maximum sunspot number SSN and the time interval between termination events ΔT, $\mathrm{SSN} = \mathcal{M}(\Delta T|\alpha,\beta) = \alpha\Delta T + \beta$, and infer the posterior density for the slope α and the intercept β of the model, $p(\alpha,\beta|\mathcal{D},\mathcal{M})$, with $\mathcal{D} = d_i = \{\mathrm{SSN}_i, \Delta T_{i-1}\}_{i=2}^{24}$ the (past) observed data in Table 1 by [4].

A particular choice of likelihood function for the special case of a straight-field model, when there are independent errors in both data coordinates, is the following (see e.g., [2])

$$\begin{aligned} p(\mathcal{D}|\mathcal{M},\alpha,\beta) \;&=\; (2\pi)^{-N/2}\left(\prod_{i=1}^{N}(\sigma^2_{\mathrm{SSN}_i} + \alpha^2\sigma^2_{\Delta T_{i-1}})^{-1/2}\right) \\ &\times\; \exp\left\{\sum_{i=1}^{N}\frac{-[d_i - \mathcal{M}(\Delta T_{i-1}|\alpha,\beta)]^2}{2(\sigma^2_{\mathrm{SSN}_i} + \alpha^2\sigma^2_{\Delta T_{i-1}})}\right\}, \end{aligned} \tag{4}$$

with each σ_{SSN} and $\sigma_{\Delta T}$ expressing the uncertainty on sunspot number and time interval between termination events, respectively.

The combination of likelihood function and uniform priors over certain ranges leads to the posterior distribution. Figure 1 shows an example solution for given fixed values for the uncertainty on the sunspot number and the time-interval between termination events. Well-constrained distributions are obtained for the marginal posteriors of the model parameters α and β (top and middle panels on the left).

Following the definition in Equation 3, the posterior predictive distribution for the future unobserved amplitude of Solar Cycle 25, $\widetilde{\mathcal{D}} = \mathrm{SSN}_{25}$, based on the time interval for the preceding termination event, ΔT_{24}, can then be computed from the inferred posterior and by considering a Gaussian likelihood for the new data as a function of the parameter vector

$$\begin{aligned} p(\mathrm{SSN}_{25}|\mathcal{M},\alpha,\beta) &= \frac{1}{\sqrt{2\pi}}\left((\sigma^2_{\mathrm{SSN}_{25}} + \alpha^2\sigma^2_{\Delta T_{24}})^{-1/2}\right) \\ &\times \exp\left\{\frac{-[\mathrm{SSN}_{25} - \mathcal{M}(\Delta T_{24}|\alpha,\beta)]^2}{2(\sigma^2_{\mathrm{SSN}_{25}} + \alpha^2\sigma^2_{\Delta T_{24}})}\right\}, \end{aligned} \tag{5}$$

with $\sigma_{\mathrm{SSN}_{25}}$ and $\sigma_{\Delta T_{24}}$ expressing the uncertainty we are willing to consider for the future sunspot number and the last time interval between termination events, respectively.

Let us assume the termination event for Solar Cycle 24 occurred in October 2021. This leads to $\Delta T_{24} = 10.72$. The obtained posterior-predictive distribution for the maximum amplitude of Cycle 25 is displayed in the bottom-left panel of Fig. 1 and shows a well-constrained posterior density. The rounded summary of the posterior predictive distribution is $\mathrm{SSN}_{25} = 191^{+11}_{-11}$, with the estimate corresponding to the median and the uncertainties given at the 68% credible interval. The main advantage of having the posterior-predictive distribution is that it is now perfectly possible and straightforward to make probability statements. For instance, according to the result displayed in Fig. 1, the probability that the maximum amplitude of Solar Cycle 25 will fall between ~ 180 and 201 is 68%, the area under the green curve covering that percentage of the full probability mass.

Another advantage of the method is that it propagates uncertainty from the observations to the inferred quantities of interest. The right panels in Fig. 1 show how varying the uncertainty about sunspot number and time interval between termination events influences the resulting posterior-predictive distribution. Uncertainty in the sunspot number affects the dispersion of the probability distribution (top-right panel). Uncertainty on the time interval between termination events affects dispersion and the location parameter of the probability distribution (middle-right panel). Adopting an approximate formula by [1] for the sunspot number error, produces a probability distribution (bottom-right panel) with a larger uncertainty and with a median, $\mathrm{SSN}_{25} = 184^{+25}_{-22}$, that is also displaced with respect to the calculation with fixed sunspot number error.

Because the method provides us with a distribution of probability among different possible sunspot-number values, it becomes possible to quantify the predictive capabilities of the precursor and/or the model. While Solar Cycle 25 is underway, we can do this exercise with past solar cycles and use the method to compute posterior-predictive distributions using the data in Table 1 by [4]. Figure 2 shows that the 95% credible intervals of the computed probability distributions cover the actually observed sunspot number in all except three cases

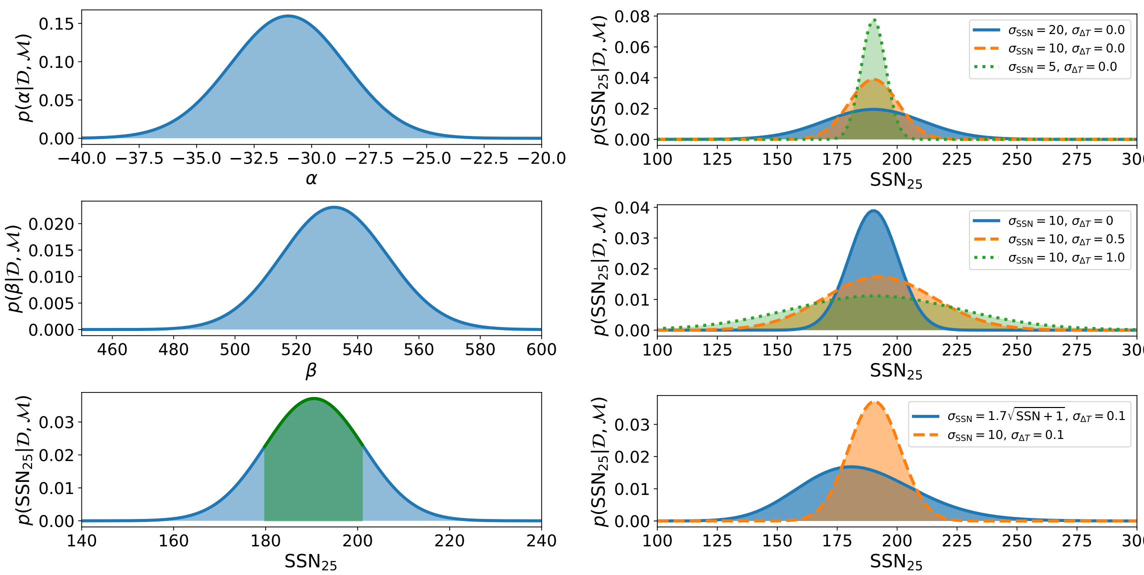

Figure 1: Left: top and middle panels show marginal posterior densities for the slope α and the intercept β of the linear model $\mathcal{M}$ that account for the past sunspot-number data $\mathcal{D}$. The calculations assume $\sigma_{\mathrm{SSN}_i} = 10$ and $\sigma_{\Delta T_{i-1}} = 0.1\,\forall i$. The bottom panel displays the posterior-predictive distribution for the sunspot-number during Solar Cycle 25, based on the posterior from past data and the likelihood of new data. The shaded green area covers 68 % of the mass centred around the median: $\mathrm{SSN}_{25} = 190.6^{+10.6}_{-10.8}$. The calculation assumes $\Delta T_{24} = 10.72$. Right: top panel shows the influence of uncertainty on sunspot-number $[\sigma_{\mathrm{SSN}}]$ on the posterior-predictive distribution for SSN_{25} for the case with no uncertainty on the time interval between termination events: $\sigma_{\Delta T} = 0$. The middle panel shows the influence of uncertainty about the time interval between termination events $[\sigma_{\Delta T}]$ on the posterior-predictive distribution for SSN_{25} for the case with $\sigma_{\mathrm{SSN}} = 10$. The bottom panel shows a comparison between posterior-predictive distributions for SSN_{25} computed with a fixed error on sunspot-number and with the approximate formula by [1].

(SC16, SC19, and SC21). The observed values are within the 68% credible interval of the prediction in 12 cases. The median of the prediction is fairly accurate in six cases. The NOAA Space Weather Prediction Center prediction interval is also shown in the figure and falls below the prediction interval for this precursor and model. On the other hand, our estimate is in good agreement with the climatological forecast by Pesnell (2018) which considers that the maximum amplitude of Solar Cycle 25 will be the average of all observed maxima.

4 Summary

In this work we suggest a method for computing predictions for the maximum amplitude of Solar Cycle 25, based on Bayesian inference, and adopting a particular precursor as example application. The relevant quantity is the posterior-predictive distribution of the maximum

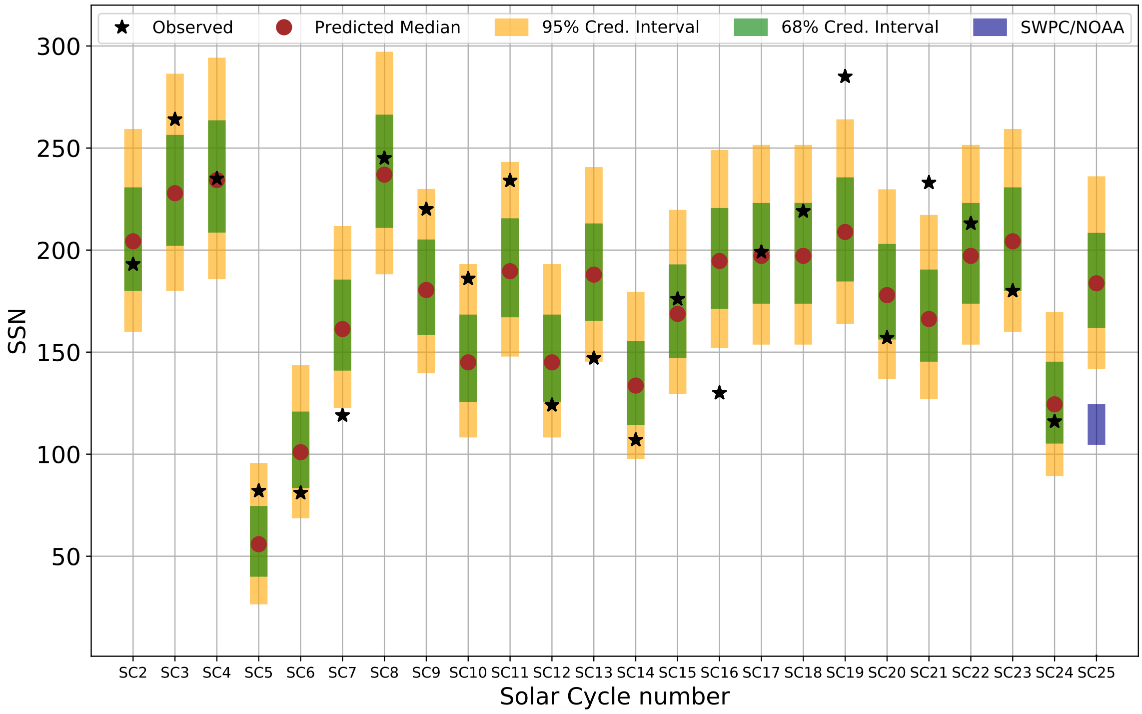

Figure 2: Comparison between observed sunspot-number values for past Solar Cycles 2 to 24 from Table 1 of [4] and posterior-predictive distributions computed using Equation (5) with data on Table 1 of [4]. In all calculations, $\sigma_{\Delta T} = 0.1$ and the approximate formula by [1], $\sigma_{\rm SSN} = 1.7\sqrt{\rm SSN + 1}$, are employed . For Solar Cycle 25, $\Delta T = 10.72$ is employed. Also shown is the interval between the minimum and maximum peak sunspot-number values currently predicted by the SWPC/NOAA. Colours indicate different credible intervals of the obtained posterior distributions.

sunspot number. It is a probability distribution and results from the combination of a posterior probability distribution, inferred from past data, and a likelihood function for unobserved future data. The useful qualities of the method are the following: it enables to make probability statements about the quantity of interest, the SSN; the inference considers the propagation of uncertainty from observables to inferred quantities; and it is applicable to other predictive methods and alternative models, hence can be used to assess the quality of the predictor and/or the adopted model.

Acknowledgements

This study was motivated by correspondence with Scott W. Mcintosh. I am grateful to Andrés Asensio Ramos, José Manuel Vaquero, and Víctor Manuel Sánchez Carrasco for helpful comments and suggestions. This research was supported by projects PGC2018-102108-B-I00 and PID2021-127487NB-I00 from Ministerio de Ciencia, Innovación y Universidades and FEDER funds.

References

[1] Dudok de Wit, T., Lefèvre, L., & Clette, F. 2016, Solar Phys., 291, 2709

[2] Gregory, P. C. 2005, Bayesian Logical Data Analysis for the Physical Sciences: A Comparative Approach with 'Mathematica' Support (Cambridge University Press)

[3] Hathaway, D. H. 2015, Living Reviews in Solar Physics, 12, 4

[4] McIntosh, S. W., Chapman, S., Leamon, R. J., Egeland, R., & Watkins, N. W. 2020, Solar Phys., 295, 163

[5] Nandy, D. 2021, Solar Physics, 296, 54

[6] Petrovay, K. 2020, Living Reviews in Solar Physics, 17, 2