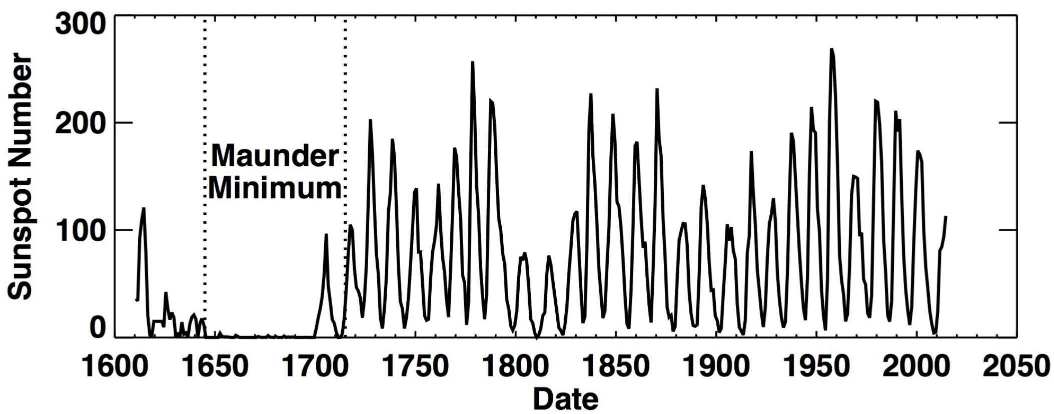
ASTP 720 - Homework 9 - Forecasting Solar Cycles

Due Date: November 24th, 2020

Solar observations have occurred for many thousands of years, including activity in the form of sunspots. Text in the Chinese *I Ching* notes a darkening within the Sun, placing the potential first observations of sunspots at earliest circa 1000 BC[1]. Deliberate observations of sunspots by Chinese astronomers trace to later times[2], with regular observational records taking place by 28 BC[1]. In Western astronomy, sunspot observations trace back to at least the Greek scholar Theophrastus in the 4th century BC[3]. Debates on the origin of sunspots occurred throughout history, with references to obscurations of some form or smaller bodies eclipsing the Sun. With the Greek belief that the Heavens were perfect, it was hard for many in the West to accept that the Sun itself could have marred in some way. Even with the advent of the telescope and the application to astronomy in the early 1600s, these debates continued, with for example Jesuit astronomer Christoph Scheiner arguing that they were small satellites orbiting the Sun[4] while Galileo Galilei believed that they had to be on our near the Sun, e.g., like clouds. It was during this time that regular observations began with telescopes, from which we can reconstruct over 400 years of sunspot observations.



Historical sunspot record, smoothed by year. Credit NASA/Marshall Solar Physics, `https://solarscience.msfc.nasa.gov/images/ssn_yearly.jpg`

Swiss astronomer Rudolf Wolf is credited with compiling the observations going back to around 1610 AD. In 1843, German astronmer Heinrich Schwabe noted an approximately 11-year periodicity in the number of sunspots[5], of which Wolf later creater a numbering scheme for the cycles, with the 1755-1766 cycle being the first. In December 2019, we started Solar cycle 25 with a new minimum. The 11-year cycleis thought to arise from the 22-year Babcock-Leighton dynamo cycle, which involves the wrapping

---

[1] Xu Zhentao 1990, "Solar Observations in Ancient China and Solar Variability," *Philosophical Transactions of the Royal Society of London*, 330, 1615, 513

[2] There is some debate, but it seems to be at least circa 200 BC if not earlier, see Stephenson 1990, "Historical Evidence concerning the Sun: Interpretation of Sunspot Records during the Telescopic and Pretelescopic Eras," *Philosophical Transactions of the Royal Society of London*, 330, 1615, 499

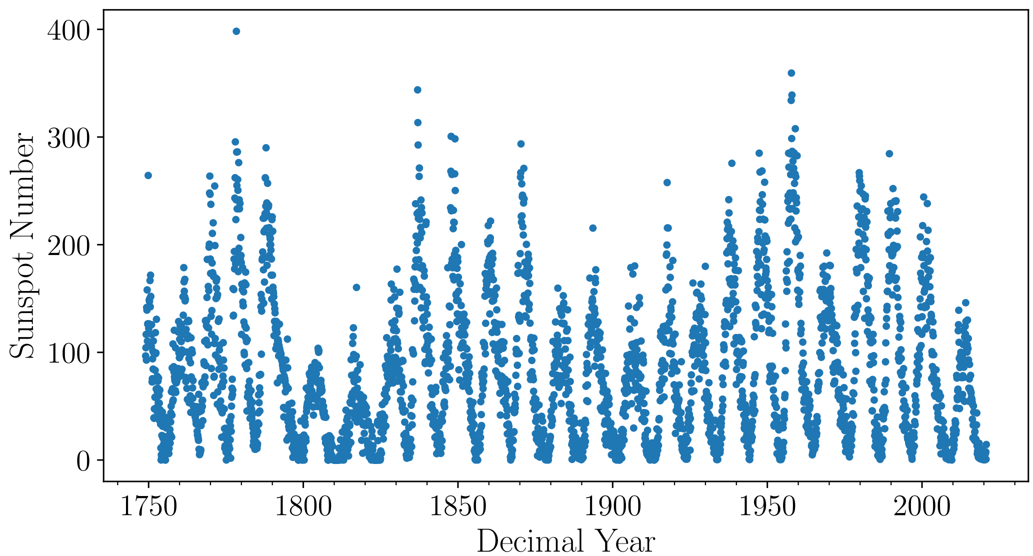
[3] Vaquero 2007, "Letter to the Editor: Sunspot observations by Theophrastus revisited ," *Journal of the British Astronomical Association*, 117, 346

[4] He argued very carefully against them being "a product of the eye, the telescope, or the air" and also did discuss the issue of irregular shapes and orientations (van Helden 1996, "Galileo and Scheiner on Sunspots: A Case Study in the Visual Language of Astronomy," *Proceedings of the American Philosophical Society*, 140, 3, 358)

[5] Schwabe 1844, "Sonnenbeobachtungen im Jahre 1843" ("Solar Observations in the Year 1843"), *Astronomische Nachrichten*, 21, 15, 233

and unwrapping of the toroidal magnetic field from differential rotation, with the polarity flipped in each half of the dynamo cycle, corresponding to a single sunspot cycle.

In `SN_m_tot_V2.0.txt`, you will find the average value of the daily sunspot number, that is the average over the given month, going back to 1749 (pre-Cycle 1). For reference, the columns are the year, month, decimal year, international sunspot number, standard deviation (not the error, the spread), and the number of observations. A -1 indicates a missing observation and an asterisk denotes a provisional value. For your work, you will really only need the sunspot number since the timeseries is equally spaced, but having the decimal year may also be useful for you. The timeseries looks like this:



You will be developing your own (seasonal) autoregressive model for the timeseries, something of the form $\phi(B)S_t = Z_t$, where $\phi(B)$ represents your model, S_t is the timeseries of sunspots, and Z_t are the usual stochastic perturbations. At a *minimum*, your model should account for:

1. the correlated nature of the month-to-month values,
2. the yearly cycle, and
3. the 11-year solar cycle.

For fitting your model, you will be using `emcee`, a common MCMC package in Python. In it, you need to define your likelihood function along with a set of priors. A snippet is provided to you so that you should first determine the residuals between the data and your model, which is equivalent to determining what the random perturbations Z_t are. Remember that in the least-squares minimization, you are trying to minimize the quantity $\sum Z_t^2$, the sum of the squared residuals/perturbations. From the residuals, you can return the log-likelihood function assuming a Gaussian. Besides fitting the parameters in your ϕ function, you may consider fitting for σ_Z, the standard deviation of the Z_t distribution (or σ_Z^2, thevariance) and you may also consider fitting for a μ, but I leave this to you.

Note that in addition to being available on myCourses, the code snippets and data file are available directly from github: `https://github.com/mtlam/ASTP-720_F2020/tree/master/HW9`

Your tasks:

[1.] In your write-up, please write down and describe your model.

[2.] Using the `emcee` package and the code snippets provided to you, fit your model to the data. You should show the fit to your model by making a corner plot with the `corner` package and write down your best-fit parameters. In addition, you should plot *both* the residuals of the fit *and* the fit of the timeseries to the data (remember that data = signal + noise). The code snippets have been provided to you either in Jupyter notebook or standard script form. The Jupyter notebook can run locally or can be uploaded to Google Colab (`https://colab.research.google.com`) if you do not wish to install `emcee` and `corner`. Please see the notebook for more information

[3.] Plot the spectrum of your model. Please describe its behavior(s), for example at low/high/peak frequencies.

[4.] Using the parameters of your model, as well as σ_Z either derived from your MCMC fit or estimated from the residuals after your fit, predict the sunspot number out to the year 2050.

[Bonus.] Recall the Akaike Information Criterion as a way of determining a preferred model, defined as

$$\mathrm{AIC} = 2k - 2\ln\mathcal{L}_{\max},$$

where k is the number of parameters of the model and $\ln\mathcal{L}_{\max}$ is the maximum value of the likelihood. Create two different models from the one you have made above. For all three, calculate the parameters and the AIC value. Determine which model is preferred and offer some interpretation of the values (see the end of the notes from Lecture 15)