**Scuola Galileiana di Studi Superiori**
**Course:** Information theory
**Academic Year:** 2022-23

Elisa Ferrero elisa.ferrero@studenti.unipd.it
Andrea Pizzolato andrea.pizzolato.2@studenti.unipd.it
Andrea Sainaghi andrea.sainaghi@studenti.unipd.it

# Predicting Solar Activity

## 1 Introduction

Our aim is to predict the influence of Jupiter on the solar cycles.
We have data that count the number of sunspots seen every month for more than the past 200 years. Using these data we want to predict the solar cycle's modes and see if there is a link with the orbital period of Jupiter. We want to achieve this result by using Bayesian inference techniques.

## 2 Theory Setup

We assume that the solar cycle is composed by $K$ modes so that each month we expect a number of sunspots given by

$$y_i = \sum_{j=1}^{K} \left[ A_j \cos(2\pi\nu_j\, t_i) + B_j \sin(2\pi\nu_j\, t_i) \right] + \sigma\eta_i \tag{1}$$

where $i = 1, ..., N$, and $\eta_i \sim N(0, 1)$ is the white noise associated and the parameters we want to find are $A_j, B_j, \nu_j, \sigma$ with $j = 1, ..., K$.
We assume a uniform prior for all the $A_j, B_j, \nu_j$ and the Jeffrey's prior for $\sigma$, namely $f(\sigma) = \sigma^{-1}$.
Since the white noise is Gaussian distributed with standard deviation $\sigma$ we have that the Likelihood function reads

$$\mathcal{L}(\vec{y}_{obs}|\vec{\theta}) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ -\frac{1}{2\sigma^2} \left[ y_{obs,i} - \sum_{j=1}^{K} \left( A_j \cos(2\pi\nu_j\, i) + B_j \sin(2\pi\nu_j\, i) \right) \right]^2 \right\} \tag{2}$$

where $\vec{\theta} = (\vec{\nu}, \vec{A}, \vec{B}, \sigma)$ is the vector parameters and $\vec{y}_{obs}$ is the vector of collected data where at each month $i = 1, ..., N$ it contains the number of observed sunspots. In our conventions $i = 1$ is January 1800 and it goes until $i = N = 2678$ which is February 2023.

We make use of Bayesian inference. Namely, given the prior and the likelihood function of the model, we want to evaluate the posterior $f(\vec{y}_{obs}|\vec{\theta})$. This approach to information theory wants to mimic the real distribution from which data are extracted by making use of the information in the already observed data. To do so we make use of Metropolis Algorithm, which is implemented in Python in the library EMCEE. This algorithm is based on simulating the steps of the model with a normal step weighted by an acceptance rate plus a term related to the jump weighted by the *rejection probability*, which is defined as

$$r(\theta') = 1 - \int t(\tilde{\theta}|\theta')a(\tilde{\theta}|\theta')d\tilde{\theta} \tag{3}$$

This object can be used to understand how the algorithm works. When we do large jumps the rejection probability is near to 1 and we have some kind of plateaus, with a small acceptance rate, whereas when there are small jumps there is a small rejection probability, and the algorithm takes a long time to explore the whole space. This means we need the right trade-off, and in many applications this means that we introduce some other type of moves.

## 3 Data Analysis

We use the EMCEE library on python to implement Metropolis algorithm. This library contains various types of sampler beyond the one we use for the Metropolis, with different parameters to tune for the family and the specific parameters of the distribution. To make the algorithm converge easily is important to have nice initial values for the parameters we need to find. To get a nice ansatz for them we proceed in the following way.

- We take the sunspot spectrum over the 200 years and we make a Fourier transform of it (after having performed a constant shift so that the overall average of $y_{obs,i}$ is null).
- We identify the mode with the greatest amplitude and we assign those values to $\nu_1$, $A_1$ and $B_1$.
- We subtract the mode just found from the sunspot spectrum and we repeat the procedure until we are left with just noise in the sunspot spectrum.

In the end we identified $K = 15$ modes, and the initial parameters fit as shown in Figure 1. In order to choose the initial value of $\sigma$ we used the empiric average of the deviation between reconstructed signal and data. It turns out to be $\sigma = 35$.

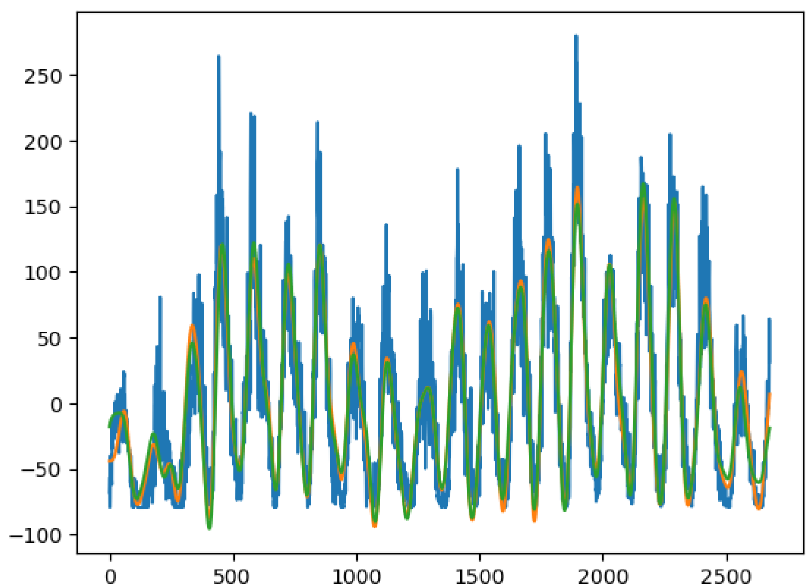


Figure 1: The blue line represents the data counting the sunspots seen every month (shifted to get null average), the green line is the reconstructed spectrum using Fourier analysis, the yellow line is the reconstructed spectrum using EMCEE.

The algorithm requires the number of walkers which is 256 in our case, the dimension of $\vec{\theta}$ which is 46 (3 parameters for each mode, two amplitudes and a frequency, and $\sigma$), the log of the posterior which is the following function[1]

$$\ln f(\vec{\theta}|\vec{y}_{obs}) = \text{const} - (N+1)\ln\sigma - \frac{1}{2\sigma^2}\sum_{i=1}^{N}\left[y_{obs,i} - \sum_{j=1}^{15}\left(A_j\cos(2\pi\nu_j\, i) + B_j\sin(2\pi\nu_j\, i)\right)\right]^2 \tag{4}$$

and finally the move used by the sampler which is "DEMove".
A plot of the final reconstructed spectrum is contained in Figure 1.
We expect that the influence of Jupiter on the solar cycles provides the biggest cycle at work. Thus we expect that the period that we want to find is associated to the biggest amplitude, namely we need to find the $j = 1, ..., 15$ such that it is maximum $A_j^2 + B_j^2$, and this is associated to the $\nu_j$ given by Jupiter.
We have that the biggest mode is found at $j = 1$ (as expected since our initial conditions are made such that the first mode was the biggest one). Using the posterior we made a plot of the distribution of $T_1 = 1/\nu_1$ and it has a gaussian-like shape. Thus we made a fit and we found that the period of the sun's cycle influenced by Jupiter is given by

$$T = (11.07 \pm 0.02)\,\text{years} \tag{5}$$

The plot of the distribution of $T_1$ with the relative fit is reported in Figure 2. The fit is good and shows a reduced $\chi^2$ of $\chi^2 = 1.71$ which not that bad.

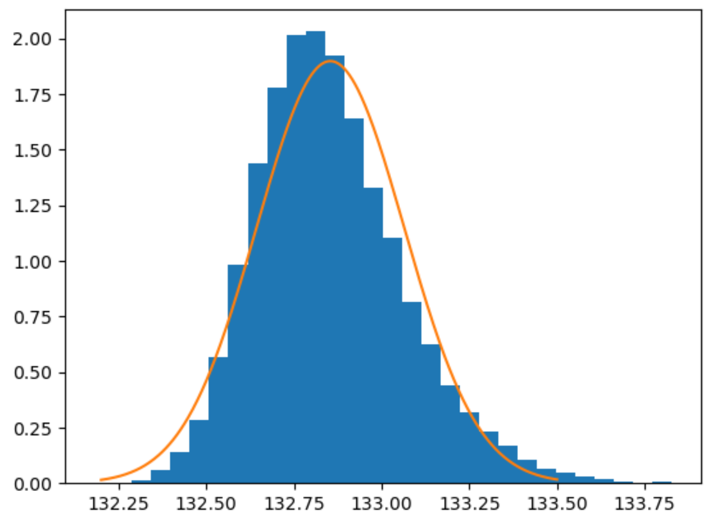


Figure 2: This is the distribution of the period $T_1$ (where on the x-axys the time is in months). The yellow line is the gaussian fit of the distribution.

# 4 Conclusion

We have analysed the number of sunspots seen every month on the sun surface to study the solar activity. Using metropolis algorithm we have shown that there exist an influence of Jupiter on the solar cycles and its effect brings a cycle of period estimated to be

$$T = (11.07 \pm 0.02)\,\text{years} \tag{6}$$

which is very similar to the revolution period of Jupiter, exactly as we hoped to get.

[1] To be precised to make the algorithm to converge easily we have used as parameters the periods of the cycles instead of the frequencies.