Link to HW9 code

1 Task 1

This homework involved using the Python *emcee* package to find the best-fitting values for a seasonal autoregressive (AR) model. Given the variation of the data on month-, year-, and 11-year-long timescales, it made sense to consider a model such as

$$X_t = \phi_1 X_{t-1} + \phi_{12} X_{t-12} + \phi_{132} X_{t-132} + Z_t \tag{1}$$

where ϕ_1 governs the month-to-month variability, ϕ_{12} governs the year-to-year variability, ϕ_{132} governs the variability of the 11-year solar cycle, and Z_t represents signal noise.

2 Task 2

The residuals (data subtracted from model) and the quality of the model compared to the data are shown in Figures 1 and 2, respectively. The residuals have a mean of 0 as expected and the fit matches the data almost perfectly.

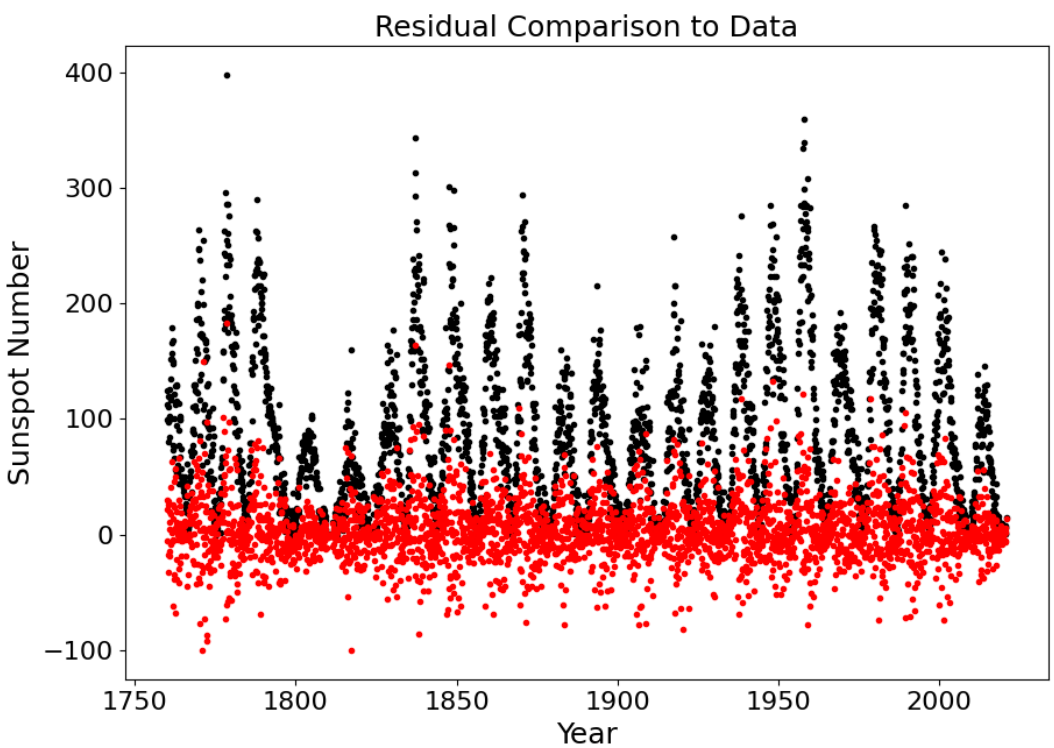

Figure 1: Plot of the residuals (data subtracted from model) for the final version of the model after 5000 iterations of which the first 25% were used for burn-in.

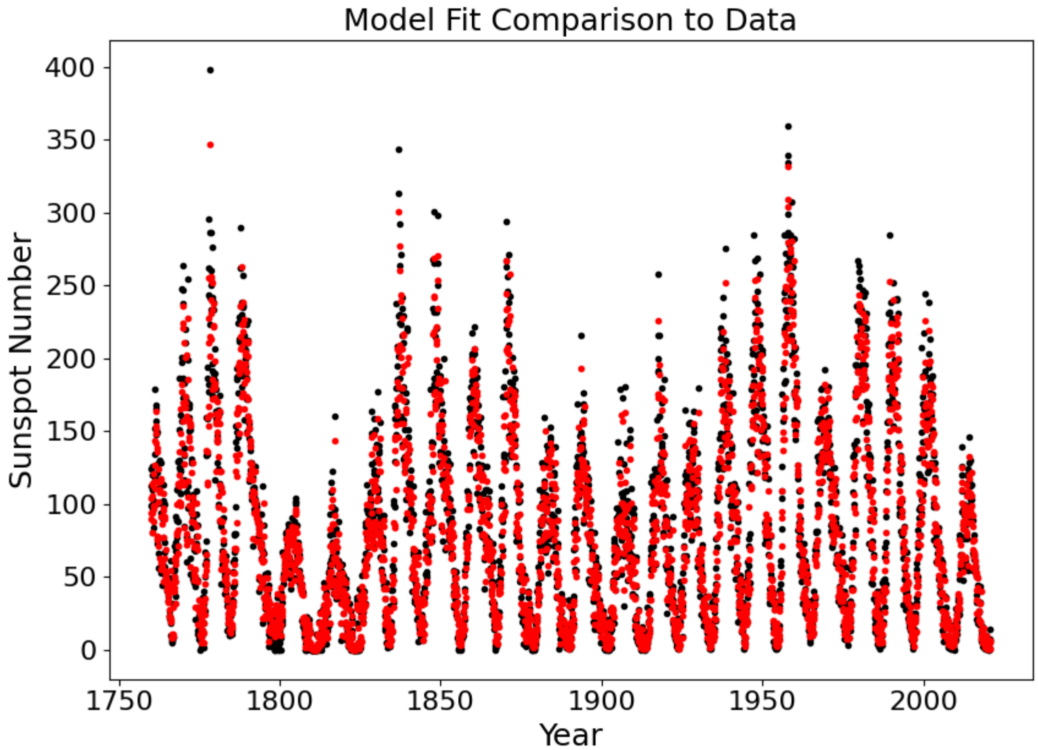

Figure 2: Comparison of the model to the data after 5000 iterations of which the first 25% were used for burn-in.

The corner plot in Figure 3 shows the posterior distributions for each of the parameters ϕ_1, ϕ_{12}, ϕ_{132}, and σ_z compared to the rest. The best-fit values after 5000 iterations of which the first 25% were used for burn-in are $\phi_1 = 0.84$, $\phi_{12} = 0.07$, $\phi_{132} = 0.05$, and $\sigma_z = 26.53$. From these values, it's clear that the monthly variations have the strongest impact on the model, whereas the year- and 11-year-long variations are comparably significant.

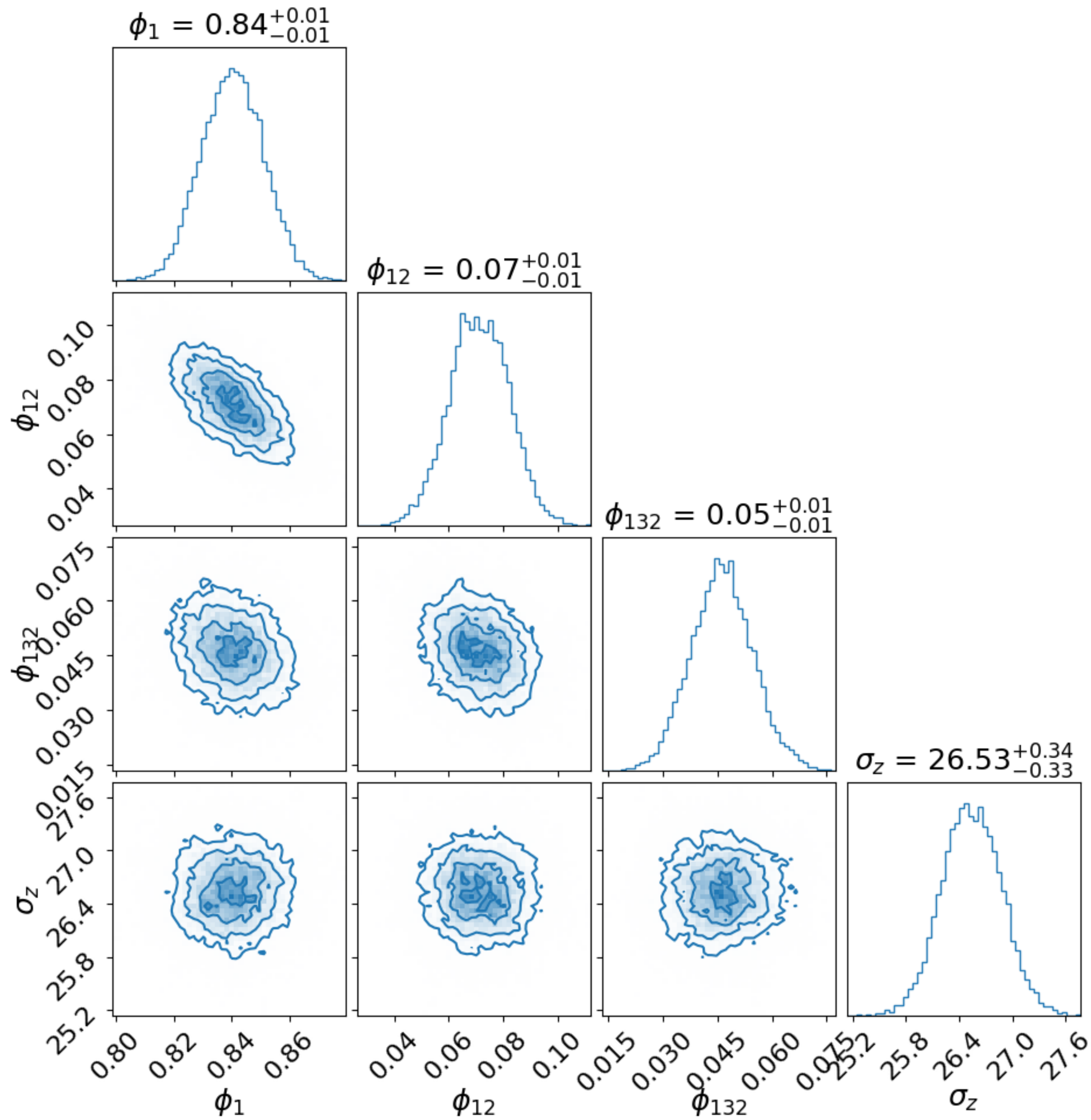

Figure 3: Corner plot showing posteriors for ϕ_1, ϕ_{12}, ϕ_{132}, and σ_z.

3 Task 3

The relevant portion of the FFT of the model is shown in Figure 4 with the left-most peak representing 52 months, the middle, second-tallest peak representing 12 months, and the right-most, most-prominent peak representing 10 months. The peaks at 10 and 52 months are strange, especially considering they are not clear harmonics of any of the other expected timescales, namely the month- and 11-year-long ones which are not represented in the FFT.

After further thought, it makes sense that the month-to-month and 11-year-cycle frequencies are not detected. The month-to-month variations are not periodic, but rather form a longer trend in the form of a 12-month periodic variation which was detected. The 11-year cycle, while periodic, was likely not as pronounced as the year-by-year fluctuations and must have been drowned out. It is also likely that the signal for the 11-year-cycle is buried in the rather wide peak near the lowest frequencies.

It is also interesting to note that the model had the month-to-month variation as the heaviest influence on the fit whereas the FFT indicated that the most obvious periodic behavior comes from the year-to-year behavior.

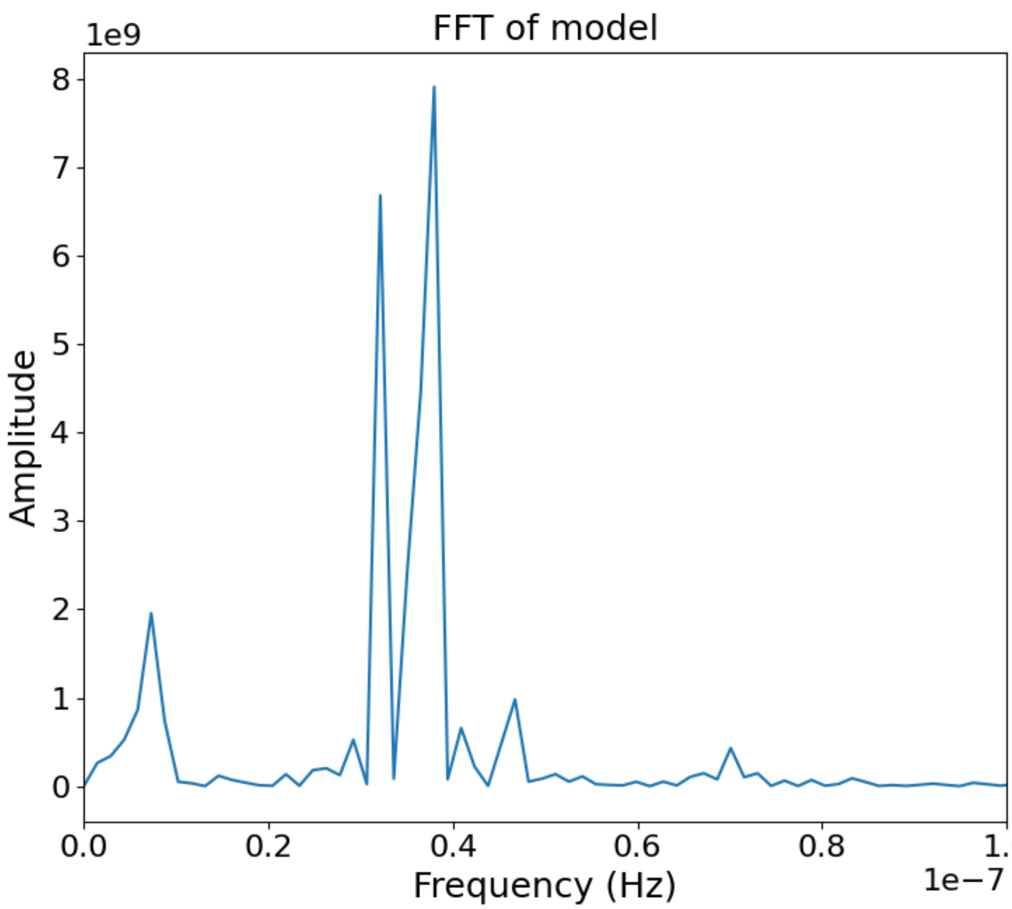

Figure 4: Relevant frequencies in the FFT corresponding to 52, 12, and 10 months in order from left to right.

4 Task 4

The model prediction out to the year 2050 is shown in Figure 5. It is obvious from the "squished" peaks that something was not correctly considered in my prediction. I tried to incorporate a σ_z factor, but depending on the given month, it would result in sometimes predicting negative sunspot values. As such, I thought it would be better to stick with something that, while wrong, retains at least some level of physicality.

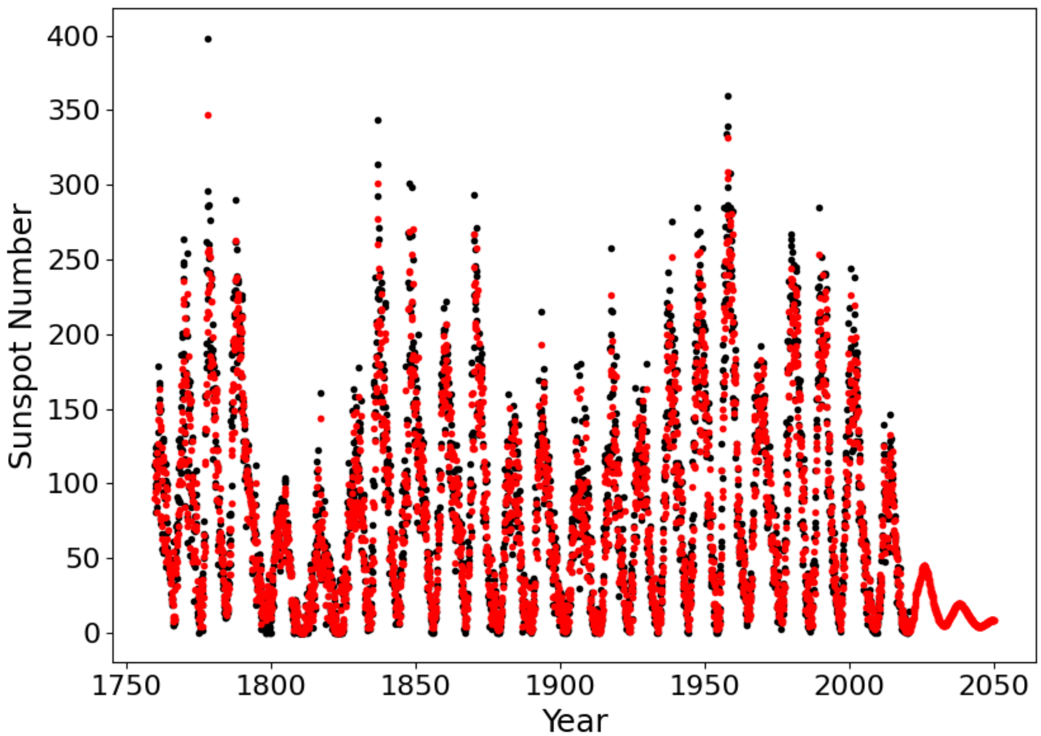

Figure 5: Prediction for sunspot number going out to the year 2050. Something is clearly not correctly considered in my prediction calculation.