Computational Methods HW- 9

Aasim Z Jan (aj3008@g.rit.edu)

November 23, 2020

https://github.com/AasimZahoor/Comp_methods.git

Question 1.

Given

Here we are given a file *SN-m-tot-V2.0* which has sunspot numbers observed at different years starting from 1749 AD. From observation it can be concluded that sunspot numbers follow two cycles, a yearly cycle and a cycle of 11 years. I have elaborated on how I approached each task and talked about if I had any issues or doubts, followed by plots/results for each task.

Approach

This task asks us to elaborate on the AR model we will be using for the tasks. The model I have chosen has μ (mean) dependence, noise (zero mean Gaussian) and has terms which account for dependence on previous value, yearly cycle and the 11 year cycle. Mathematically, my model can be written as;

$$X_t = c + \phi_1 X_{t-1} + \phi_{12} X_{t-12} + \phi_{132} X_{t-132} + N(0, \sigma_z)$$

Here, c accounts for mean (μ), ϕ_1 accounts for dependency on previous value, ϕ_{12} accounts for the yearly cycle, ϕ_{132} accounts for the 11 year cycle and N accounts for noise.

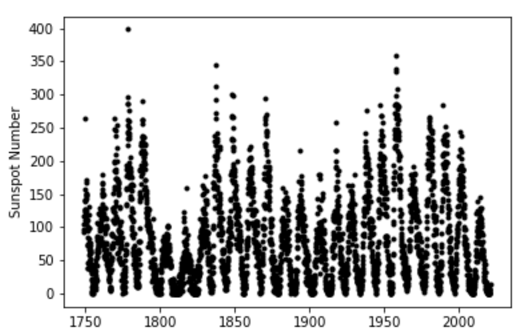
$$\text{AR model} = c + \phi_1 X_{t-1} + \phi_{12} X_{t-12} + \phi_{132} X_{t-132} + N(0, \sigma_z)$$

$$\text{Model} = c + \phi_1 X_{t-1} + \phi_{12} X_{t-12} + \phi_{132} X_{t-132}$$

$$\text{Noise} = N(0, \sigma_z)$$

I had initially though to take 132 phis to account for the 11 year cycle, that way I would have had 134 parameters. But I realized the future value has dependence on past value, the value a year back and the value 11 years back. The values in between should not determine what value it takes. So I decided against it and chose to take three phis, $\phi_1, \phi_{12}, \phi_{132}$

Plots



This plot shows the number of sunspots observed each year, staring from 1749 AD and ending at October 2020.

Question 2.

This task requires us to find the parameter values for our model using MCMC. To do that I had to first define three function to give the probability of the model with given parameter values. So basically you give the probability function the parameter values and the data and it returns the probability of the model having those parameter values. I used the same likelihood function we used in Hw_7. The likelihood function is:

$$L = \frac{1}{\sqrt{2\pi\sigma^2}} exp^{-\frac{1}{2}(\frac{data-model}{\sigma})^2}$$

The three functions are:

- **lnprior(theta)**:

  This function returns prior for the given values of theta. My function favors phi values between [-1,1]. My code ran into problems because of np.inf so I had to change it from what you had given in code you had given.

  Parameters-
  theta : Array of parameters.

  Returns-
  Value of log-prior

- **lnlike(theta,data)**:

  This function returns loglikelihood for given parameters of the model. The model is also coded in this part of the code.

  Parameters-
  theta : Array of parameters
  data : data

  Returns-
  loglikelihood of the model with the given parameters.

- **lnprob(theta, data)**:
  This function gives the probability of the model with given parameters. This function calls the previous two functions and calculates the probability.

  Parameters-
  theta : Array of parameters
  data : data

  Returns-
  Probability of the model with the given parameters.

So after defining the functions I used the emcee package to find the parameter values. I used 10 walkers, 5000 iterations, 25 percent of the parameter values were burnt. The initial values given were: were:

$$p(0) = [1, 0.5, 0.5, 0.5, 0.1]$$

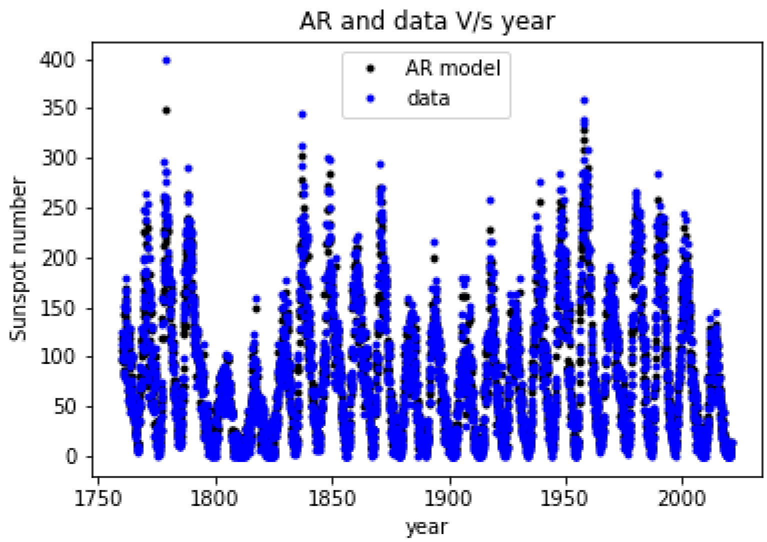
Then I took the average of the parameter values the MCMC had gone through (after burning 25 percent) to find the parameter values. Using those parameter values I found the sunspot numbers using just model and AR model. I also stored the noise used for each sunspot number. The equations I used are:

$$\text{Model} = c + \phi_1 X_{t-1} + \phi_{12} X_{t-12} + \phi_{132} X_{t-132}$$

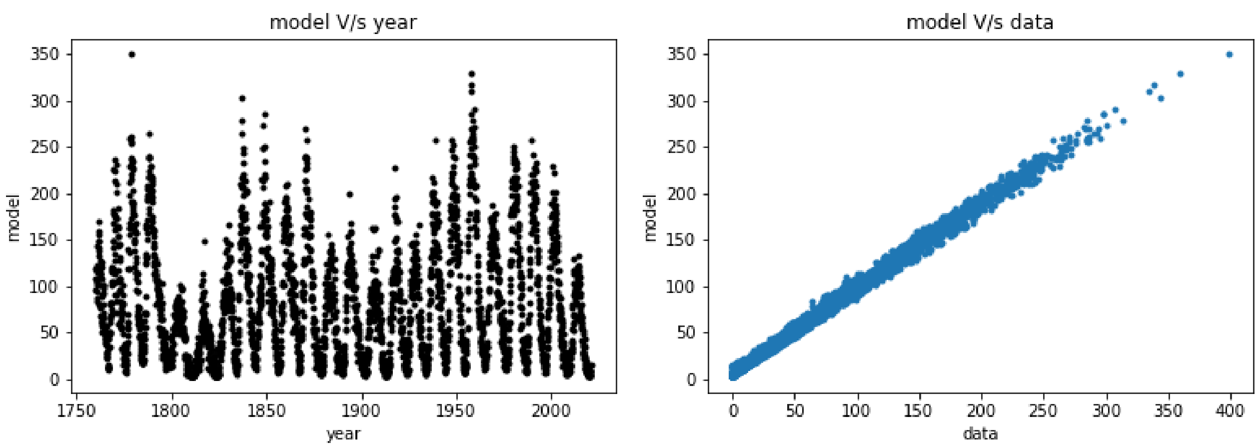
$$\text{Noise} = N(0, \sigma_z)$$

$$\text{AR model} = c + \phi_1 X_{t-1} + \phi_{12} X_{t-12} + \phi_{132} X_{t-132} + N(0, \sigma_z)$$

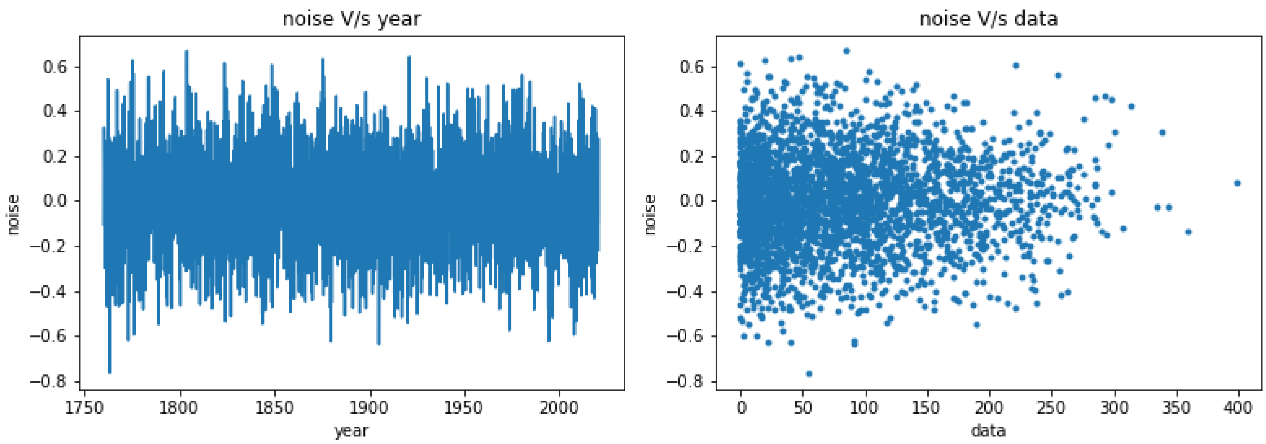
Plots for question 2



This plot shows the values of sunspot number obtained using the AR model and compares it with the data.



These plot on left is Model V/s year and the plot on right is model v/s data. A good model will have a diagonal plot and the model here has an almost diagonal plot showing it is a good model.

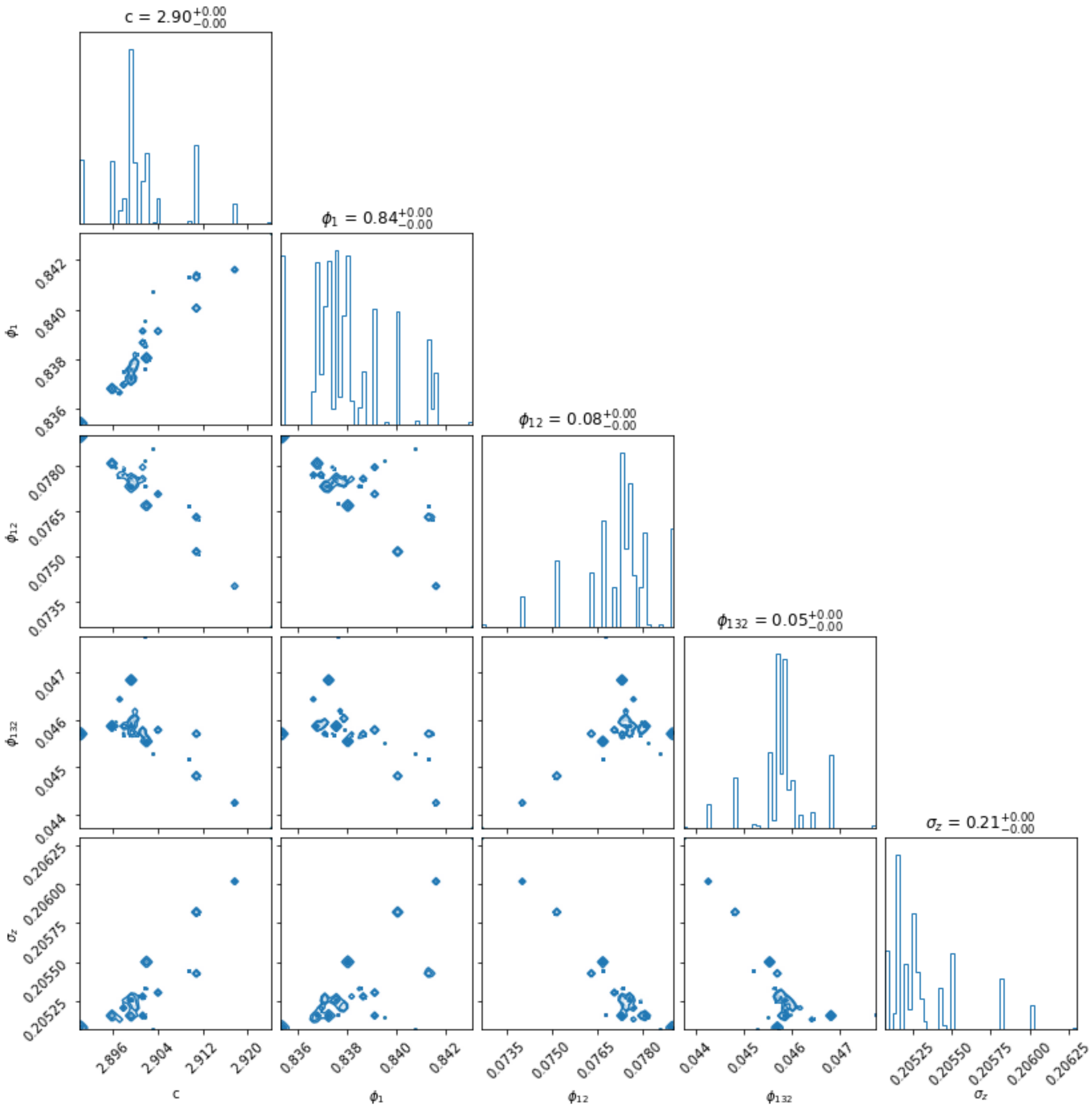


These plots show the variation of noise with year(left) and data (right).

```
In [45]: runcell('----------------------------------------plotting the model +noise
values----------------------------------', '/Users/aj3008/Desktop/MS_3rd_Sem/
Comp_methods_in_AST/Comp_methods/Hw_9/pb1.py')
c= 2.9006567104175267 phi= 0.8379331669059935 phi12= 0.0772461010580822 phi132=
0.0457817515774435 sigmaz= 0.20531355848041535
<Figure size 432x288 with 0 Axes>

In [46]:
```

This plot shows the values of the AR model parameters



This plot is the corner plot generated by the corner package for the given parameters. The average values are given at the top of each column of plots.

Question 3.

Here we are asked to plot the spectrum of our model. Initially I just thought it meant plotting the Fourier transform but after reading through my notes and lecture notes 23 I realized it means to find the Fourier transform of co variance. From the internet (link) I found for AR(p) model $S(f)$ is given as:

$$S(f) = \frac{\sigma_z^2}{|1 - \Sigma_{k=1}^{p} \phi_k e^{-2\pi jfk}|^2}$$

So initially I used the same process of finding the frequency range as we do for FFT. That gave really weird results so after searching a bit I stumbled on this link (page 7) and it says we can use any range of frequency for AR spectral analysis and that gave the plot which is in the plot section.

- **frequ()**:
  This gives the frequencies from 10^{-10} to 10^2.

  Parameters:
  none

  Returns:
  frequency in Hertz.

Earlier approach

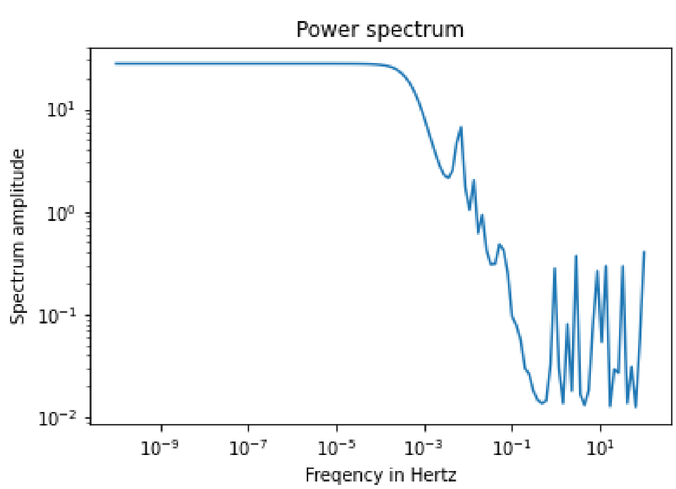
For the Fourier transform plot, I used np.fft.fft to find the Fourier transform. Using the Fourier transform I calculated the Amplitude. The frequency was calculated using a frequency function described below:

- **freq(time)**:
  This finds the frequency for the Fourier transform using the time at which data was taken.

  Parameters:
  time : time at which each sunspot value was taken(in days).

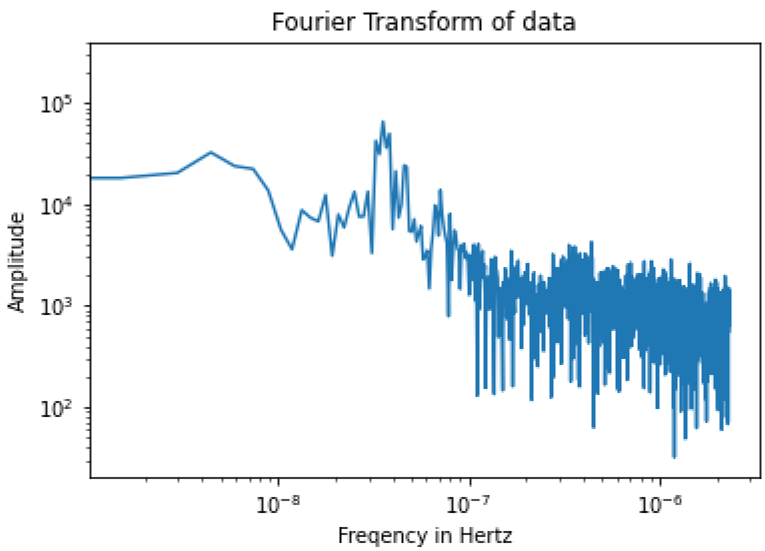
  Returns:
  frequency in Hertz.

Plots of question-3



I think the higher frequency peaks represents the 1 year cycle and the lower frequency peaks represents the 11 year cycle. I am imagining it as a wave of 1 year period enveloped by a 11 year period wave. Since the 1 year wave will have higher frequency the high frequency peaks will represent that cycle and since 11 year cycle will have relatively lower frequency, low frequency peaks will represents the 11 year. It is also observed that the lower frequencies have high amplitudes. Maybe this is because it covers 11 years, meaning a lot of values and hence has higher "power" where as high frequency covers only 1 year. This is just based what I would think it means.

Earlier approach plot-question 3



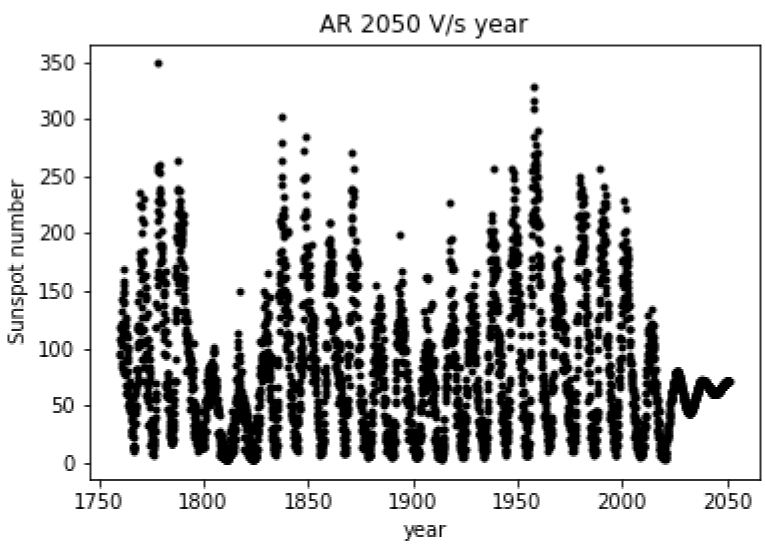
This is the plot of Fourier transform of the data.

Question 4.

We are asked to plot till year 2050. To do that I first made a new time array which goes from 1760 - 2050 (I didn't use the first 11 year since I needed them for initial values). Then using the below equation and parameter values found earlier I found the sunspot values at different values. I feel my model is not good for predicting this far in time.

$$X_t = c + \phi_1 X_{t-1} + \phi_{12} X_{t-12} + \phi_{132} X_{t-132} + N(0, \sigma_z)$$

Plot



5. Bonus

This problem asks us to find AIC for three models and hence compare the 3 models, including the model I used for earlier tasks. The AIC values were obtained using the equation;

$$\mathrm{AIC} = 2k - 2\ln(L_{max})$$

I had to make changes to the likelihood function to account for different models. So depending on how many parameters the likelihood function receives, the likelihood function will choose a model accordingly. This was achieved using if statements.

Here, I used the following three AR models. One which would account for the past value dependency, 1 and 11 year cycle, one which would account for the past value dependency and 1 year cycle, one which would account just for the past value dependency. Here k has a base value of 2 so k value will be 2 + number of parameters. Mathematically;

$$\textbf{Model-1} = c + \phi_1 X_{t-1} + \phi_{12} X_{t-12} + \phi_{132} X_{t-132} + N(0, \sigma_z)$$

This model accounts for previous value dependence, yearly cycle and 11 year cycle. k value is 7.

$$\textbf{Model-2} = c + \phi_1 X_{t-1} + \phi_{12} X_{t-12} + N(0, \sigma_z)$$

This model accounts for previous value dependence and yearly cycle. k value is 6.

$$\textbf{Model-3} = c + \phi_1 X_{t-1} + N(0, \sigma_z)$$

This model accounts for jusr previous value dependence. k value is 5.

```
In [66]: runcell(2, '/Users/aj3008/Desktop/MS_3rd_Sem/Comp_methods_in_AST/Comp_methods/
Hw_9/Bonus.py')
model 1 and 2 comparison 0.46897167345053764
model 1 and 3 comparison 0.44443400640662195

In [67]:
```

Comparison was done by taking the exponent of Δ AIC/AIC values. I made changes to this expression because my likelihood function has scaling problems (I think), so to make more sense of AIC values I made these changes.

I think it makes sense that model 1 is the best since it accounts for all dependencies and cycles. Model 2 accounts for yearly cycle so it is somewhat better than model 3.