ASTR596 Final Presentation

Group Members:
Frank Fu, Jennifer Li, Sihan Li, Chris Tandoi

Analysis & Prediction of Sunspot Numbers

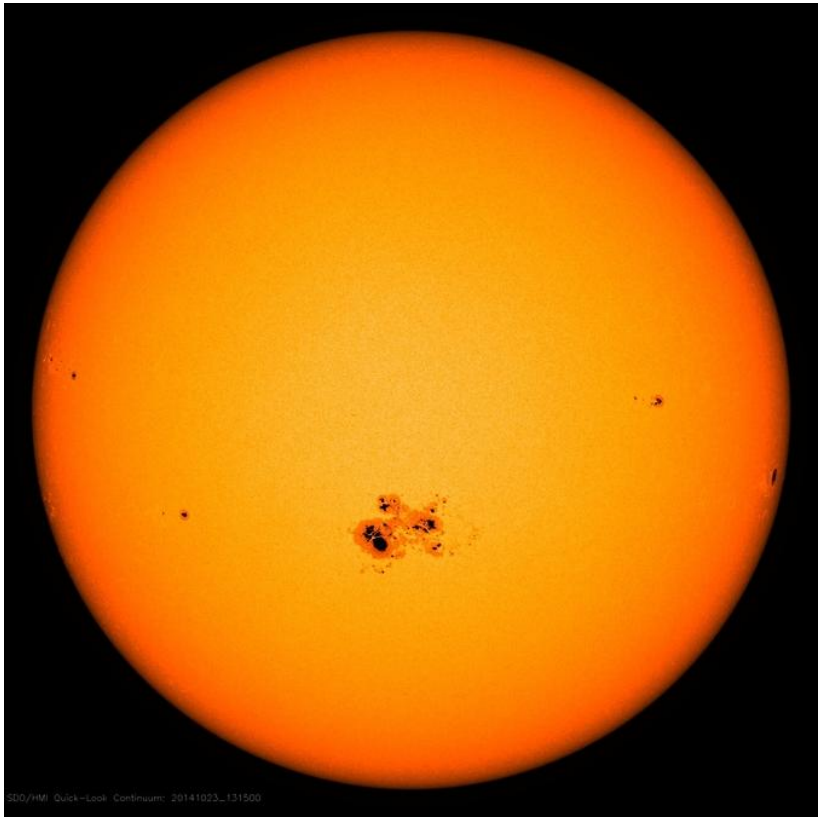
Introduction

Things to talk about:

- Physics and brief history of sunspots
- Time series analysis
- Thesis work?
- Machine Learning

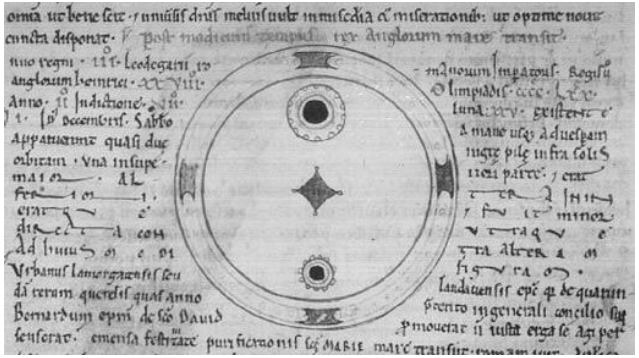
Sunspots

- Areas of particularly strong magnetic forces
- They are cooler than other part of the Sun’s surface → appear darker
- Periods of activity on the Sun known as solar cycles
- Sunspot numbers correlated with solar maximum and minimum



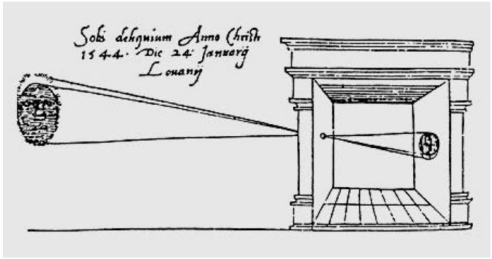
History of detection

- First recorded observations circa 800 BC in the Chinese *Book of Changes*



The Chronicle of John of Worcester - 1128 AD

Undated drawing from Tiānyuán Yùlì Xiángyìfù, manuscript 305-257 at Naikaku Bunko, Books of Shoheizaka Gakumonjo, in the National Archives of Japan [in Chinese], involved in an imperial manual of Chinese astro-omenological divination compiled in 1424–1425 (adapted from Hayakawa et al. 2017b)



Drawing of camera obscura by Gemma Frisius in 1545

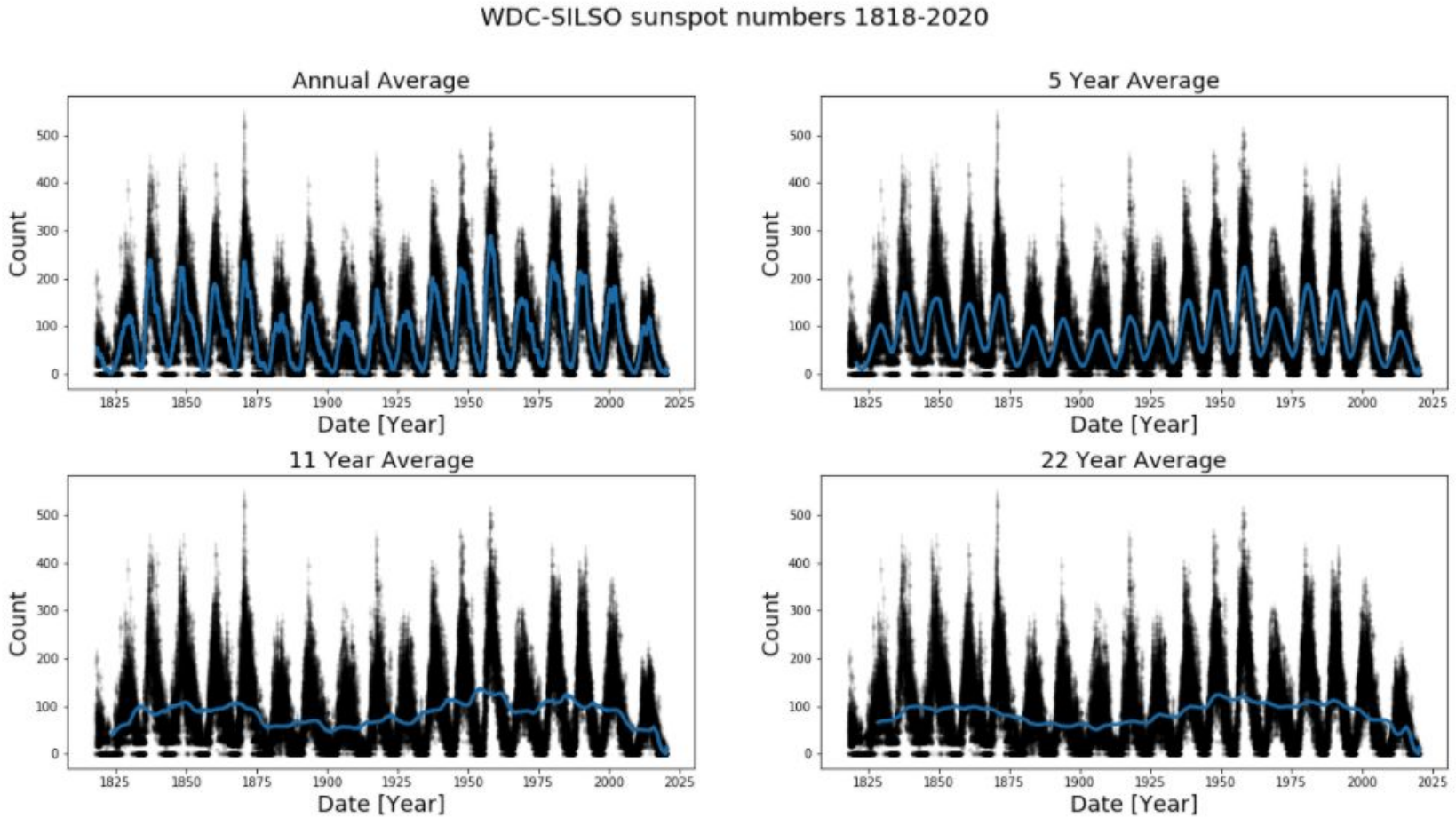
A General Map of the World, or Terraqueous Globe with all the New Discoveries and Marginal Delineations, Containing the Most Interesting Particulars in the Solar, Starry and Mundane System

Samuel Dunn - 1794

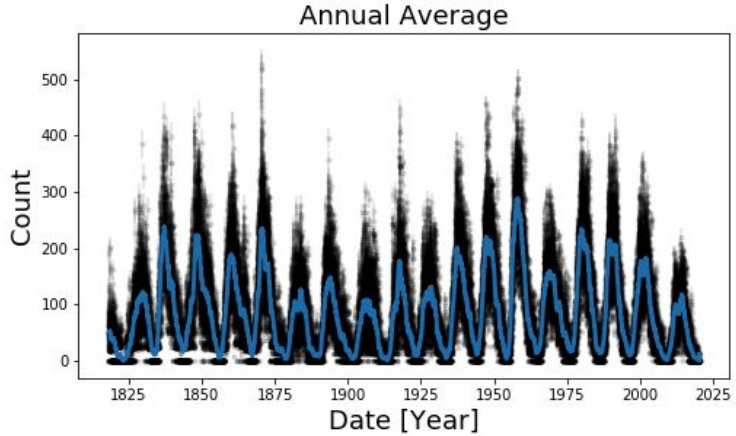
Time Series Analysis

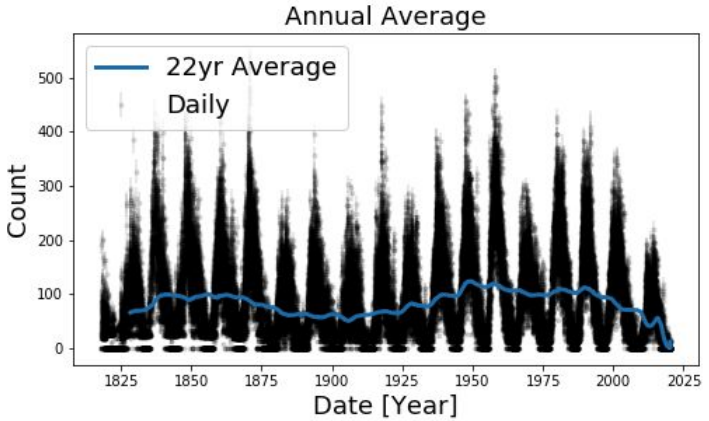
- Visualize our data
- Look at some averages
- Autocorrelation Function
- What can we do with this?

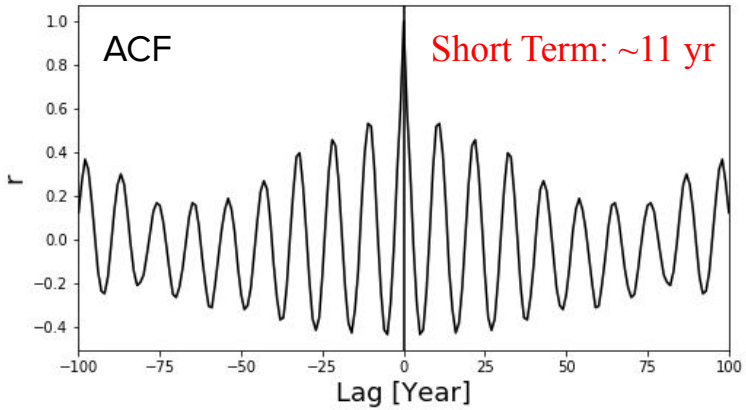
LOOK AT THE DATA

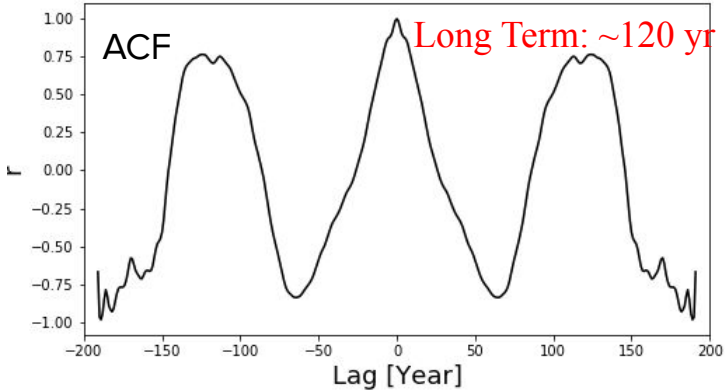


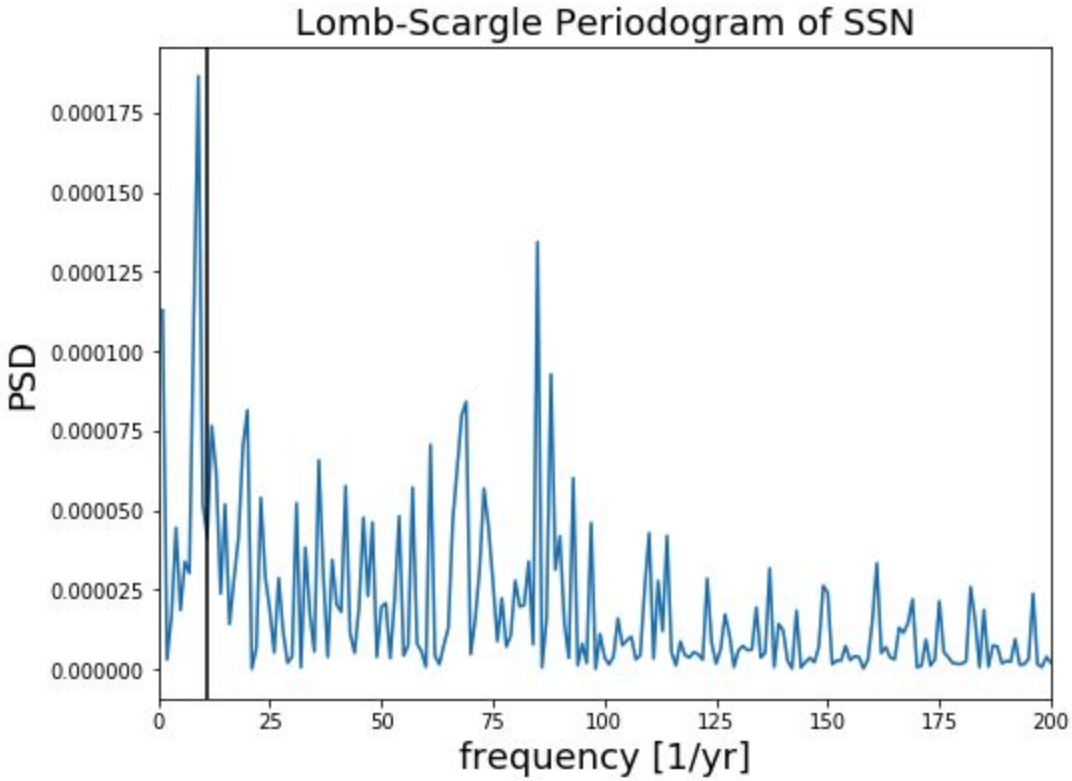
Periodicity in Sunspot Numbers







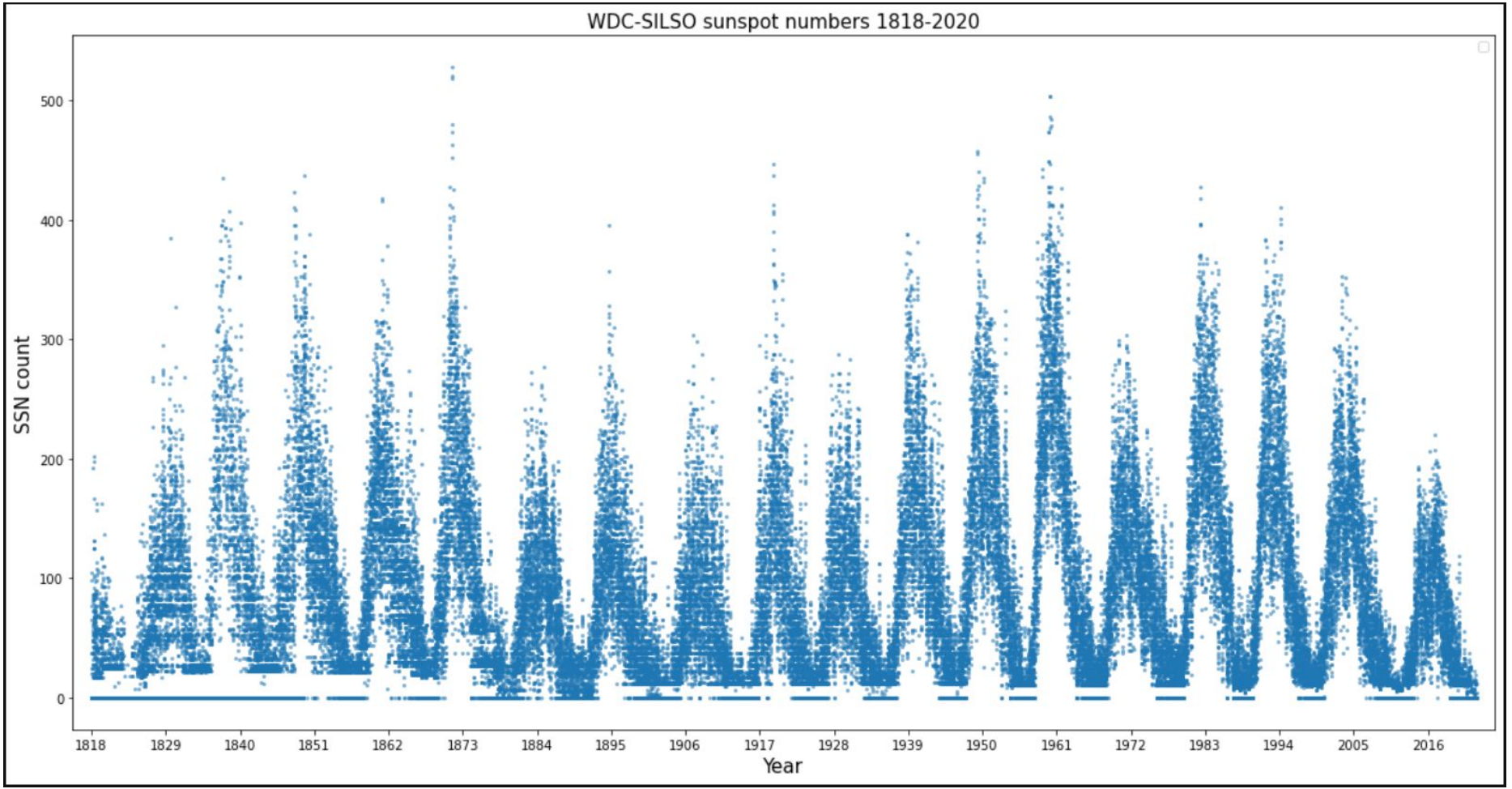


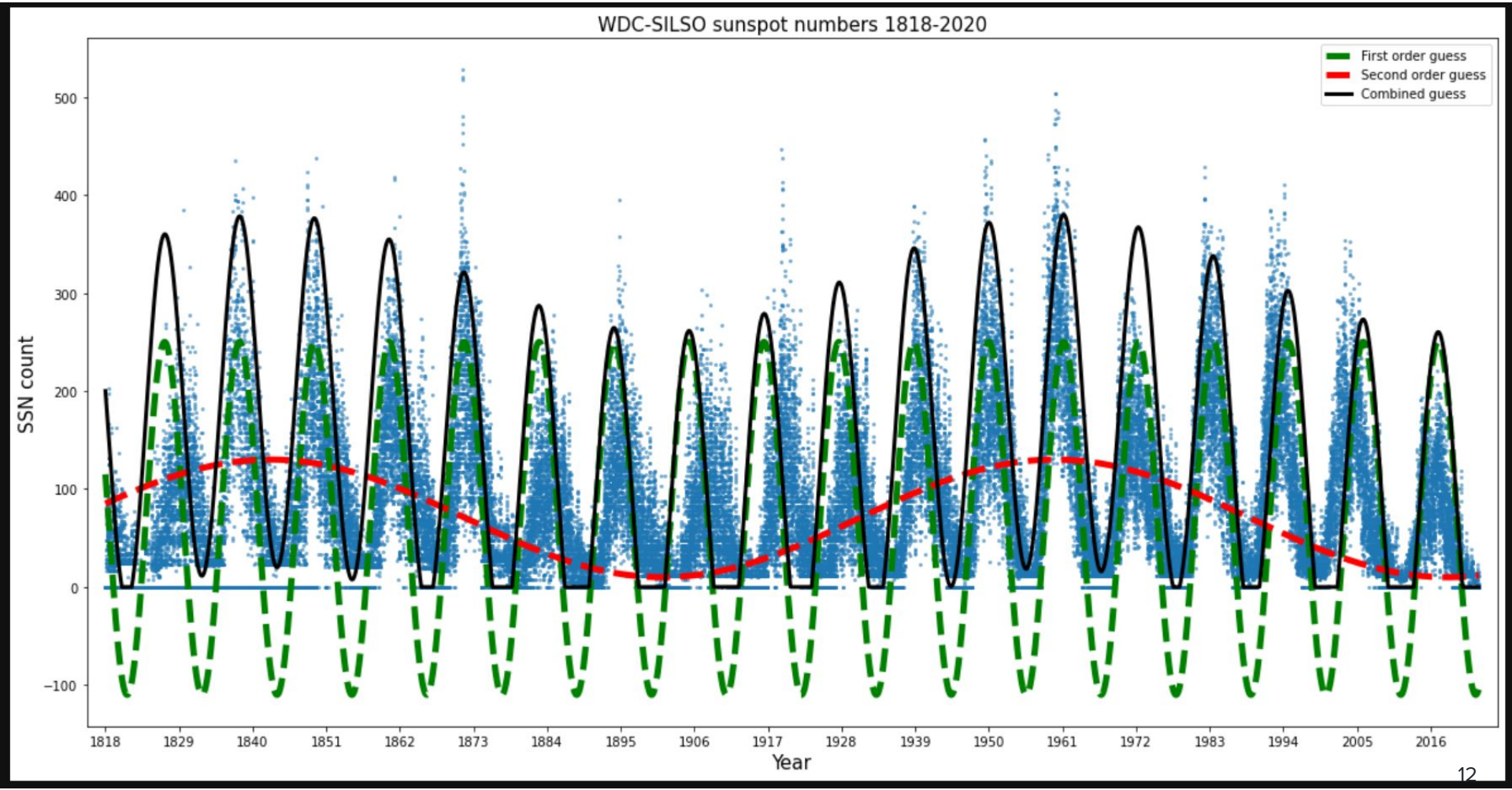

Lomb-Scargle Periodogram of SSN
PSD
0.000175
0.000150
0.000125
0.000100
0.000075
0.000050
0.000025
0.000000
0
25
50
75
100
125
150
175
200
frequency [1/yr]

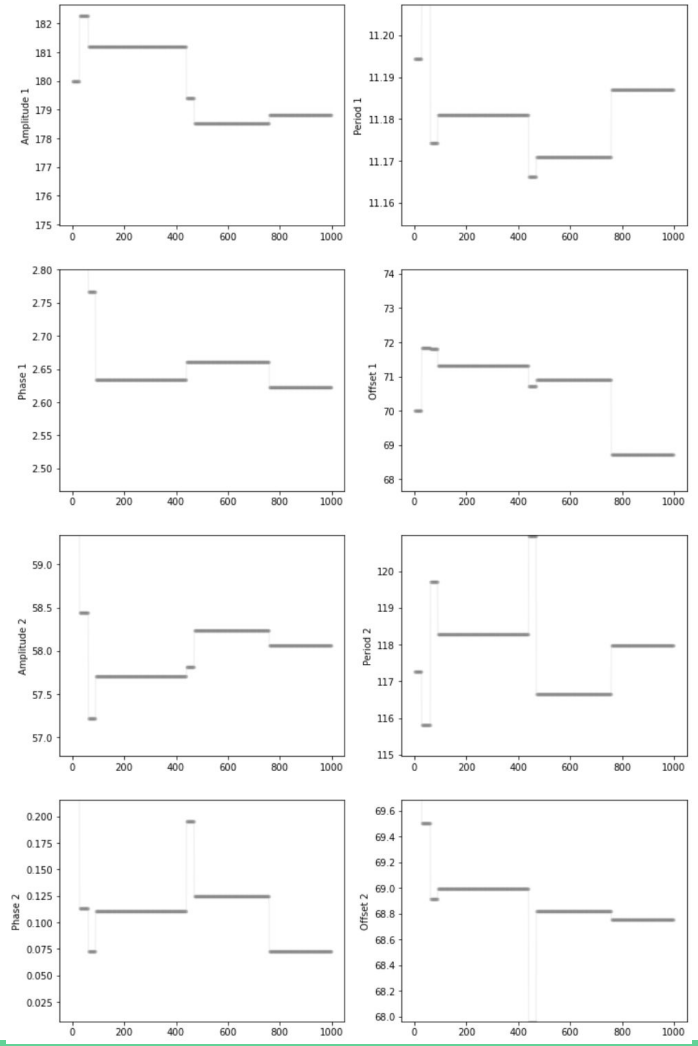
Modeling and Forecasting

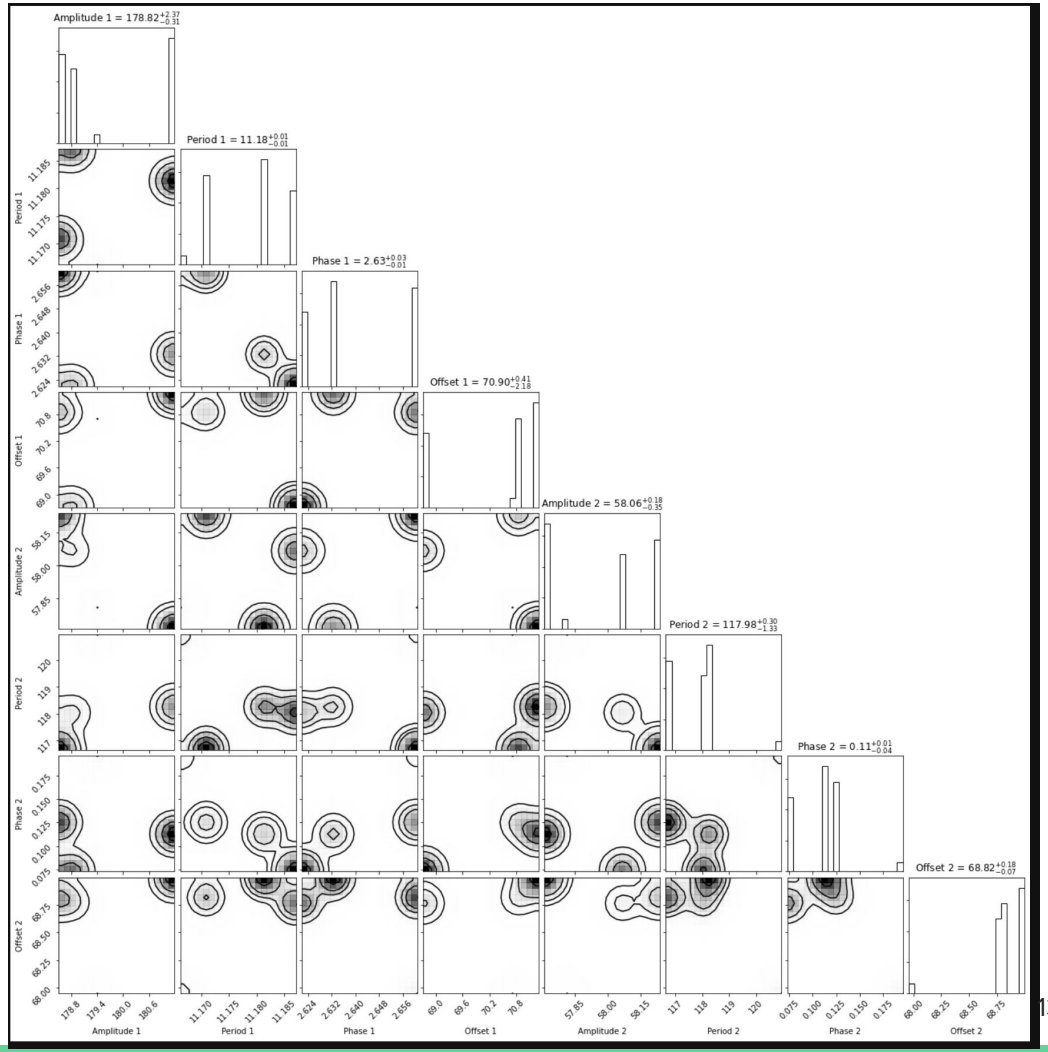
- Metropolis-Hastings
- Fourier Component Analysis
- Gaussian Process Regression
- Machine Learning Analysis

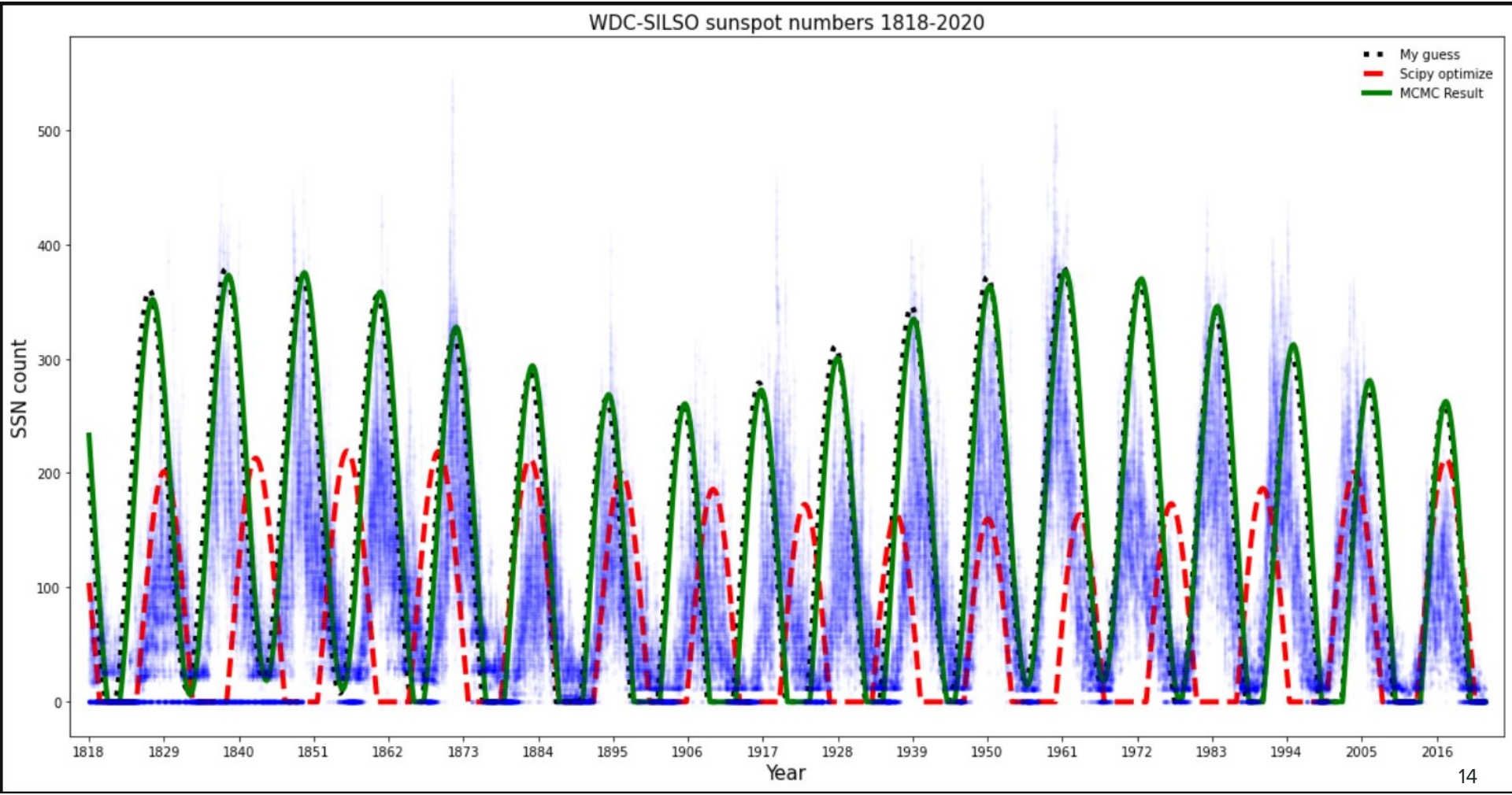
Metropolis-Hastings

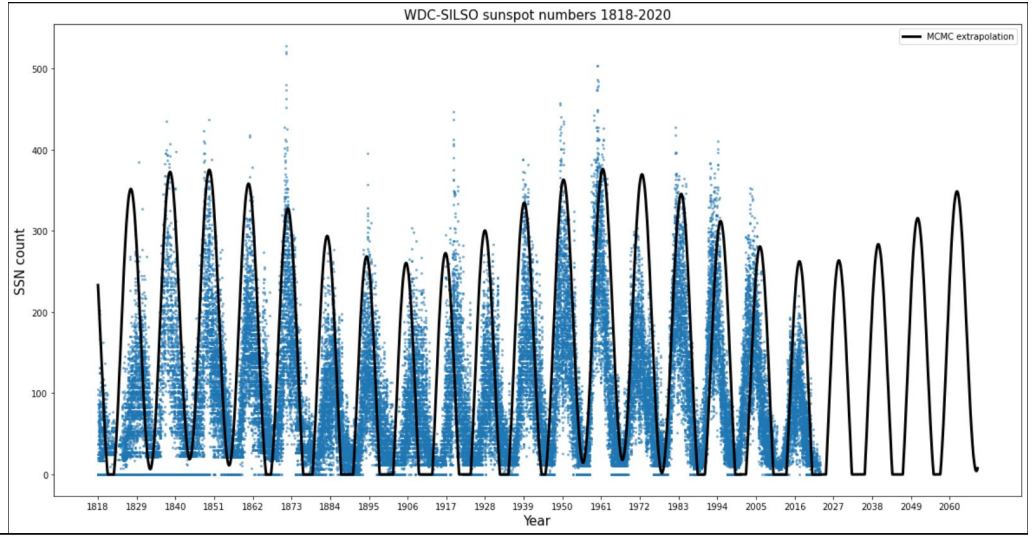

WDC-SILSO sunspot numbers 1818-2020
SSN count
0
100
200
300
400
500
1818
1829
1840
1851
1862
1873
1884
1895
1906
1917
1928
1939
1950
1961
1972
1983
1994
2005
2016
Year

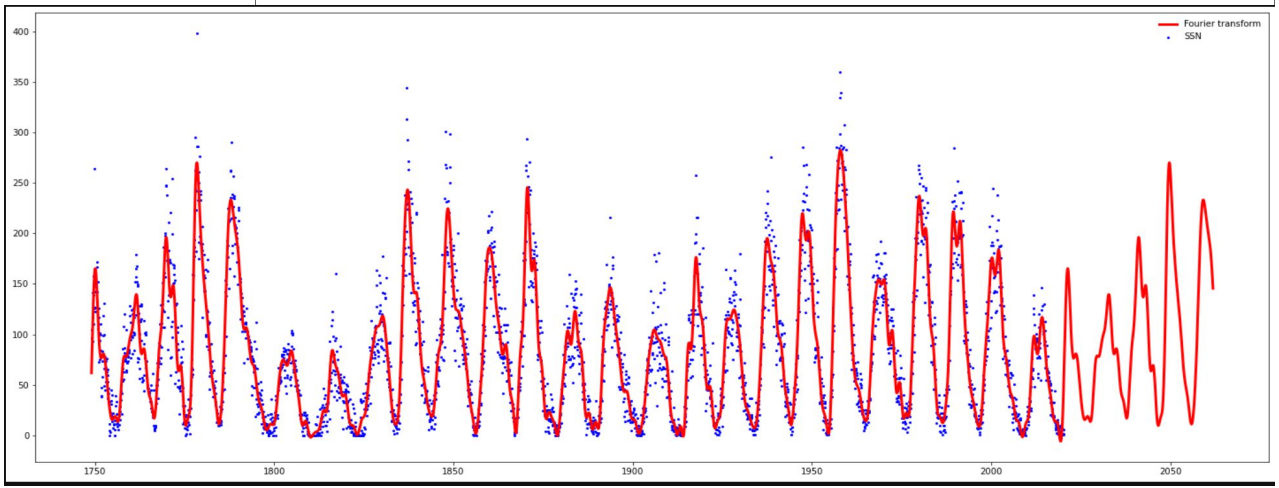

WDC-SILSO sunspot numbers 1818-2020
First order guess
Second order guess
Combined guess
SSN count
500
400
300
200
100
0
−100
1818
1829
1840
1851
1862
1873
1884
1895
1906
1917
1928
1939
1950
1961
1972
1983
1994
2005
2016
Year


Amplitude 1
Period 1
Phase 1
Offset 1
Amplitude 2
Period 2
Phase 2
Offset 2


Amplitude 1 = 178.82$^{+2.37}_{-0.31}$
Period 1 = 11.18$^{+0.01}_{-0.01}$
Phase 1 = 2.63$^{+0.03}_{-0.01}$
Offset 1 = 70.90$^{+0.41}_{-2.18}$
Amplitude 2 = 58.06$^{+0.18}_{-0.35}$
Period 2 = 117.98$^{+0.30}_{-1.33}$
Phase 2 = 0.11$^{+0.01}_{-0.04}$
Offset 2 = 68.82$^{+0.18}_{-0.07}$

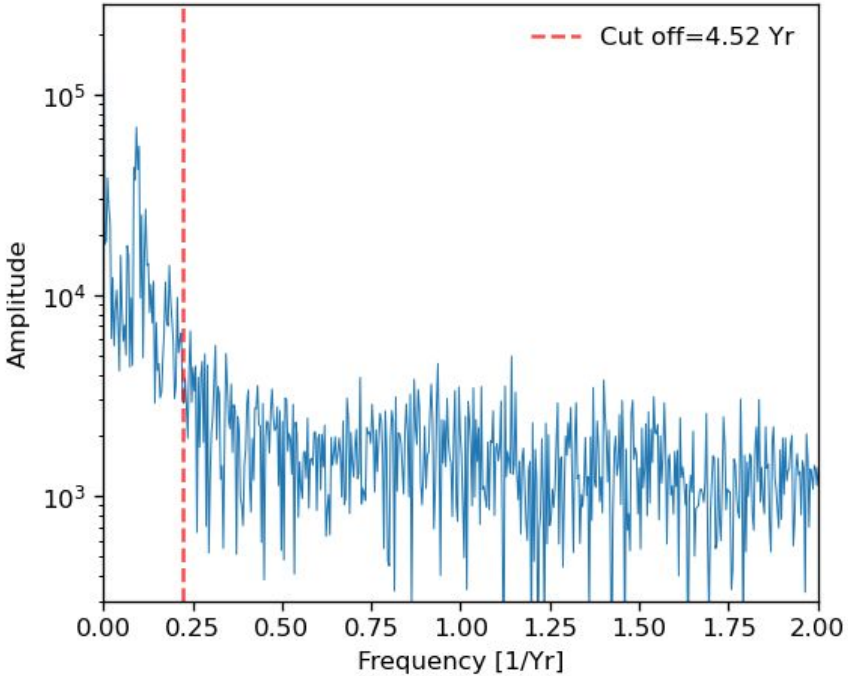

WDC-SILSO sunspot numbers 1818-2020
My guess
Scipy optimize
MCMC Result
SSN count
0
100
200
300
400
500
Year
1818
1829
1840
1851
1862
1873
1884
1895
1906
1917
1928
1939
1950
1961
1972
1983
1994
2005
2016

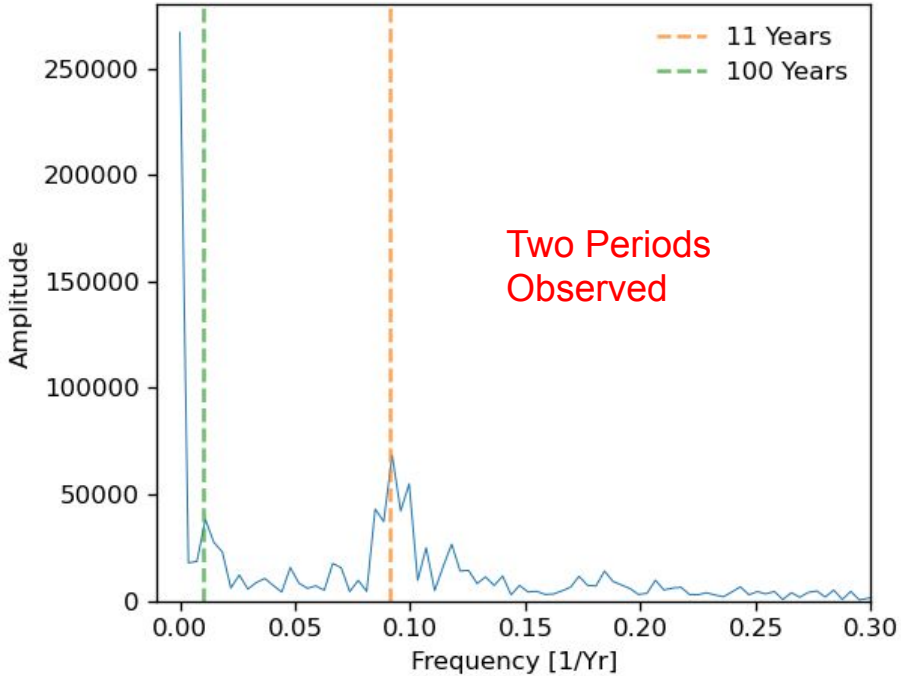

WDC-SILSO sunspot numbers 1818-2020
MCMC extrapolation
SSN count
Year
0
100
200
300
400
500
1818
1829
1840
1851
1862
1873
1884
1895
1906
1917
1928
1939
1950
1961
1972
1983
1994
2005
2016
2027
2038
2049
2060


Fourier transform
SSN
0
50
100
150
200
250
300
350
400
1750
1800
1850
1900
1950
2000
2050

Fourier Component Analysis

Power Spectral Density of SILSO monthly data

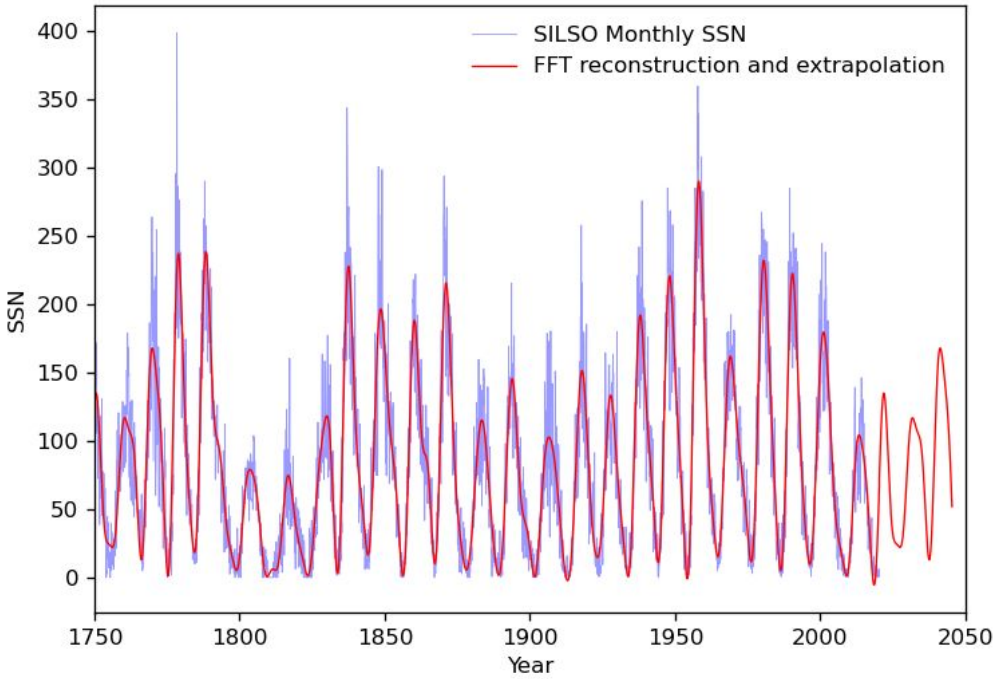


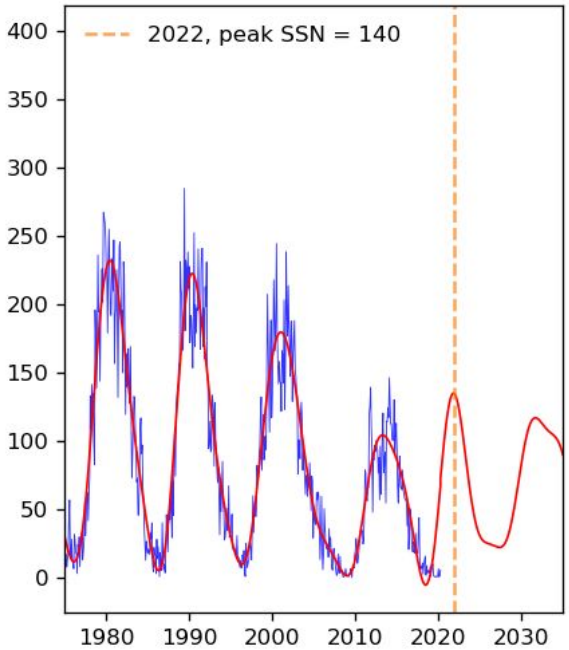


Two Periods Observed

Low Frequency reconstruction of the data

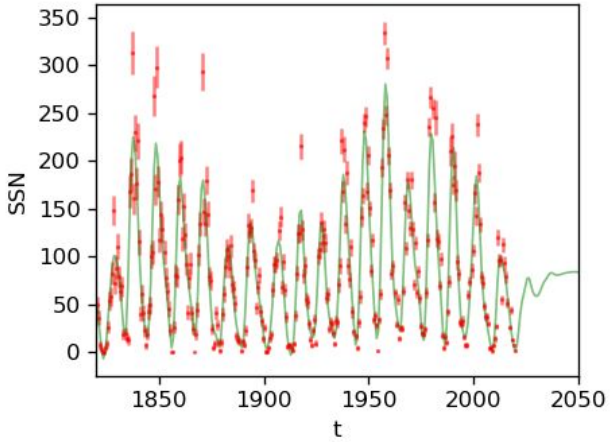
An early next peak

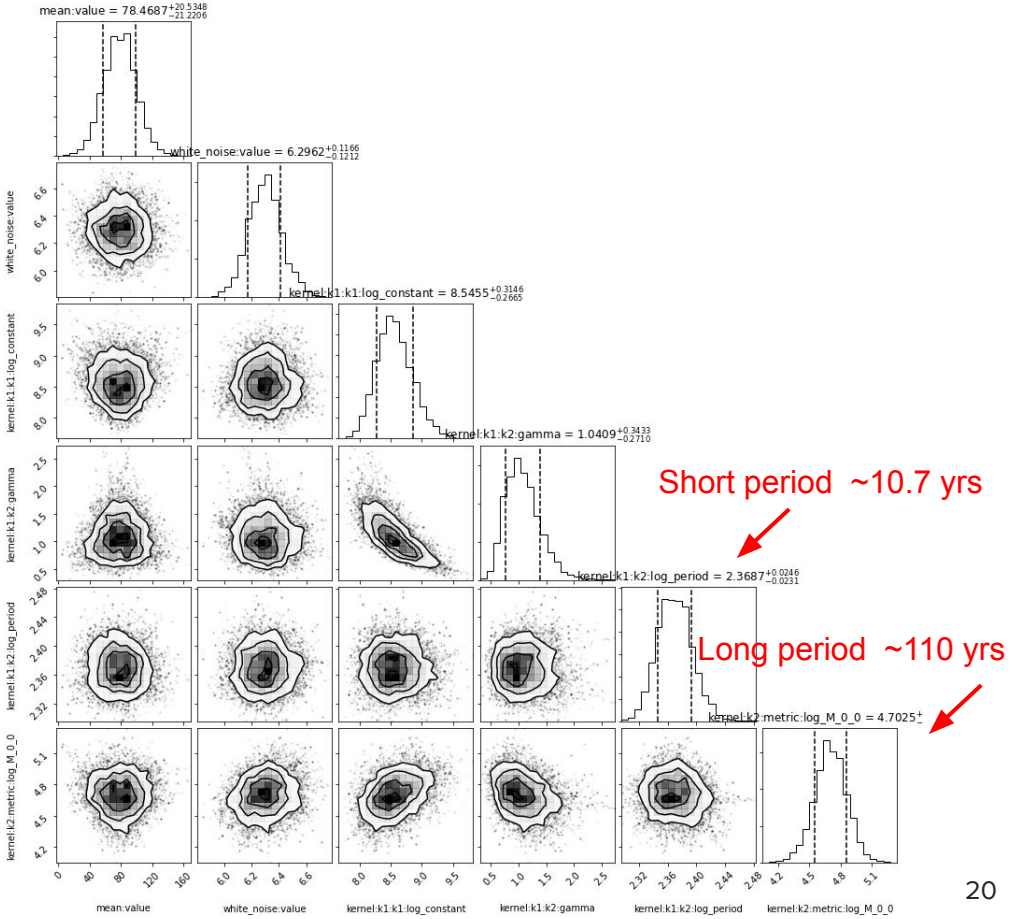




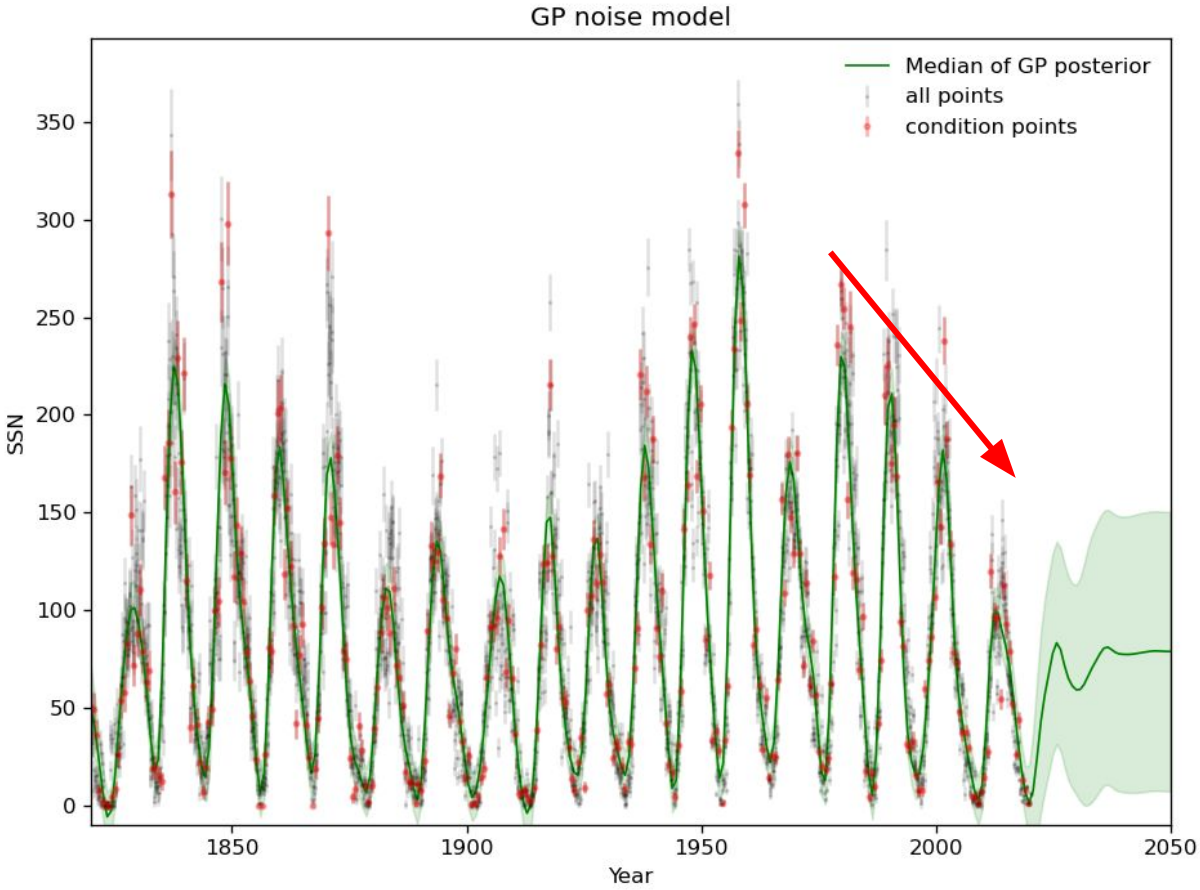
Gaussian Process Regression

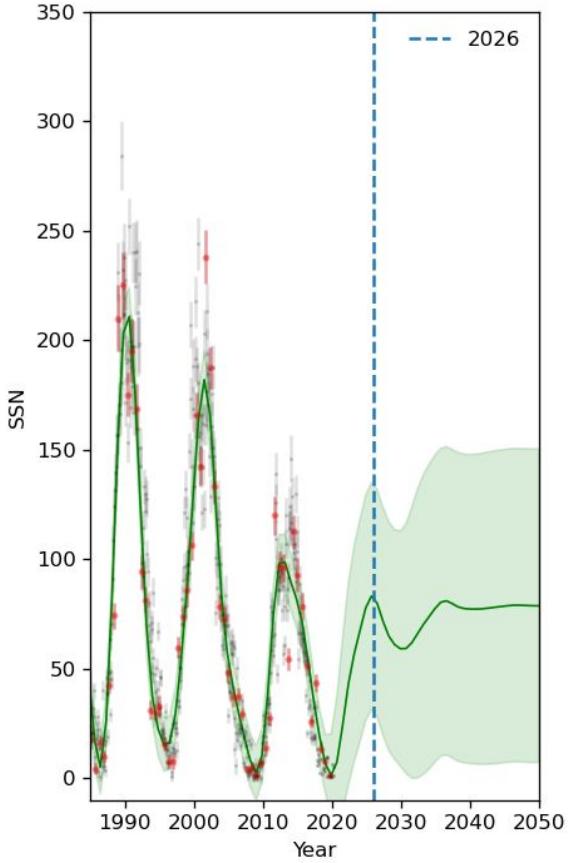
- Taking every 8th data from SILSO Monthly data
- The kernel used in GPR is an exp-sine-squared kernel combined with an exp-squared kernel
- White noise is also included in the model



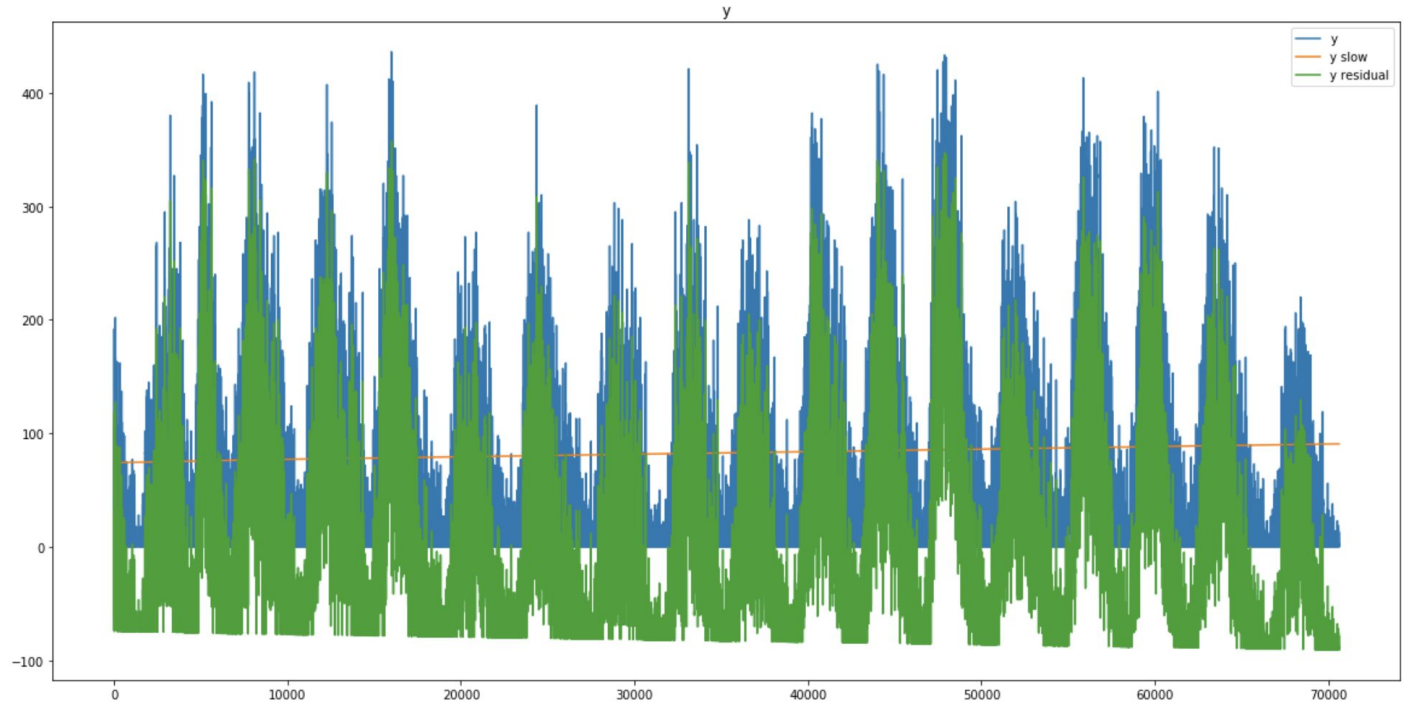


Next peak will be around 2025, with a peak SSN of 80.





Machine Learning Analysis




```
# of cycles=18
3923 days/cycle
frequency =0.00025
```

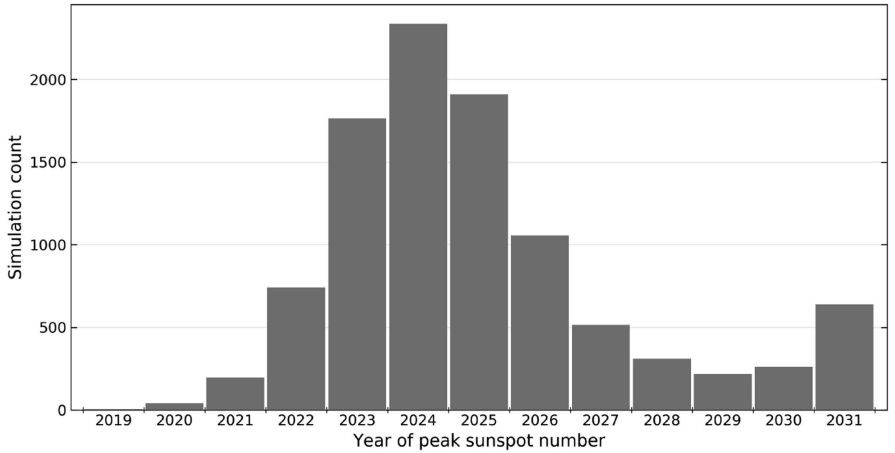
```
Accuracy of KNN classifier on training set: 0.97
Accuracy of KNN classifier on test set: 0.91
Accuracy of RF classifier on training set: 0.17
Accuracy of RF classifier on test set: 0.17
Accuracy of LDA classifier on training set: 0.16
Accuracy of LDA classifier on test set: 0.16
Accuracy of KNN classifier on training set w/ 30 neighbors: 0.64
Accuracy of KNN classifier on test set w/ 30 neighbors: 0.58
```

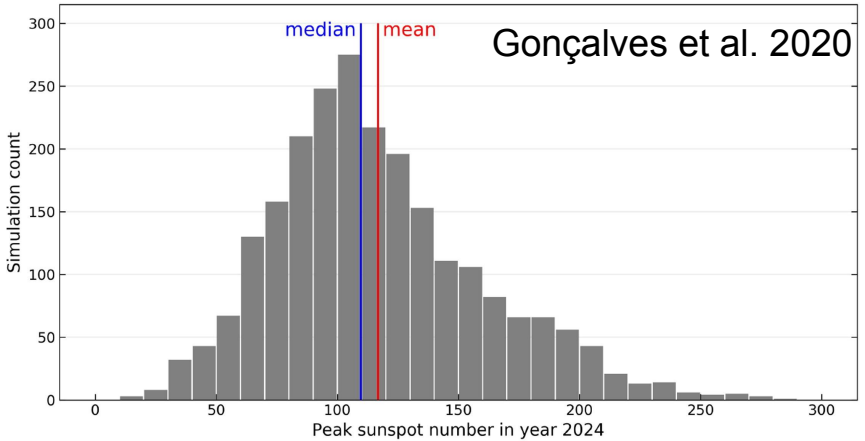
```
---------------------------------------------------------------------------
ValueError                                Traceback (most recent call last)
<ipython-input-22-052f7c73bc86> in <module>
      8 print(y_score_knn)
      9 print(y_score_knn.shape)
---> 10 fpr_knn, tpr_knn, thresholds = roc_curve(y_test, y_score_knn, pos_label=lb)
     11
     12 y_score_rf = rf.predict_proba(X_test)[:,1]

~/opt/anaconda3/lib/python3.7/site-packages/sklearn/metrics/_ranking.py in roc_curve(y_true, y_score, pos_label, sa
mple_weight, drop_intermediate)
    769     """
    770     fps, tps, thresholds = _binary_clf_curve(
--> 771         y_true, y_score, pos_label=pos_label, sample_weight=sample_weight)
    772
    773     # Attempt to drop thresholds corresponding to points in between and

~/opt/anaconda3/lib/python3.7/site-packages/sklearn/metrics/_ranking.py in _binary_clf_curve(y_true, y_score, pos_l
abel, sample_weight)
    534     if not (y_type == "binary" or
    535             (y_type == "multiclass" and pos_label is not None)):
--> 536         raise ValueError("{0} format is not supported".format(y_type))
    537
    538     check_consistent_length(y_true, y_score, sample_weight)

ValueError: multilabel-indicator format is not supported
```

Predictions

A comparison of all the predictions

Method	Next Peak	Peak SSN
Metropolis-Hastings	2028	280
Fourier Component	2022	140
Gaussian Process	2026	80

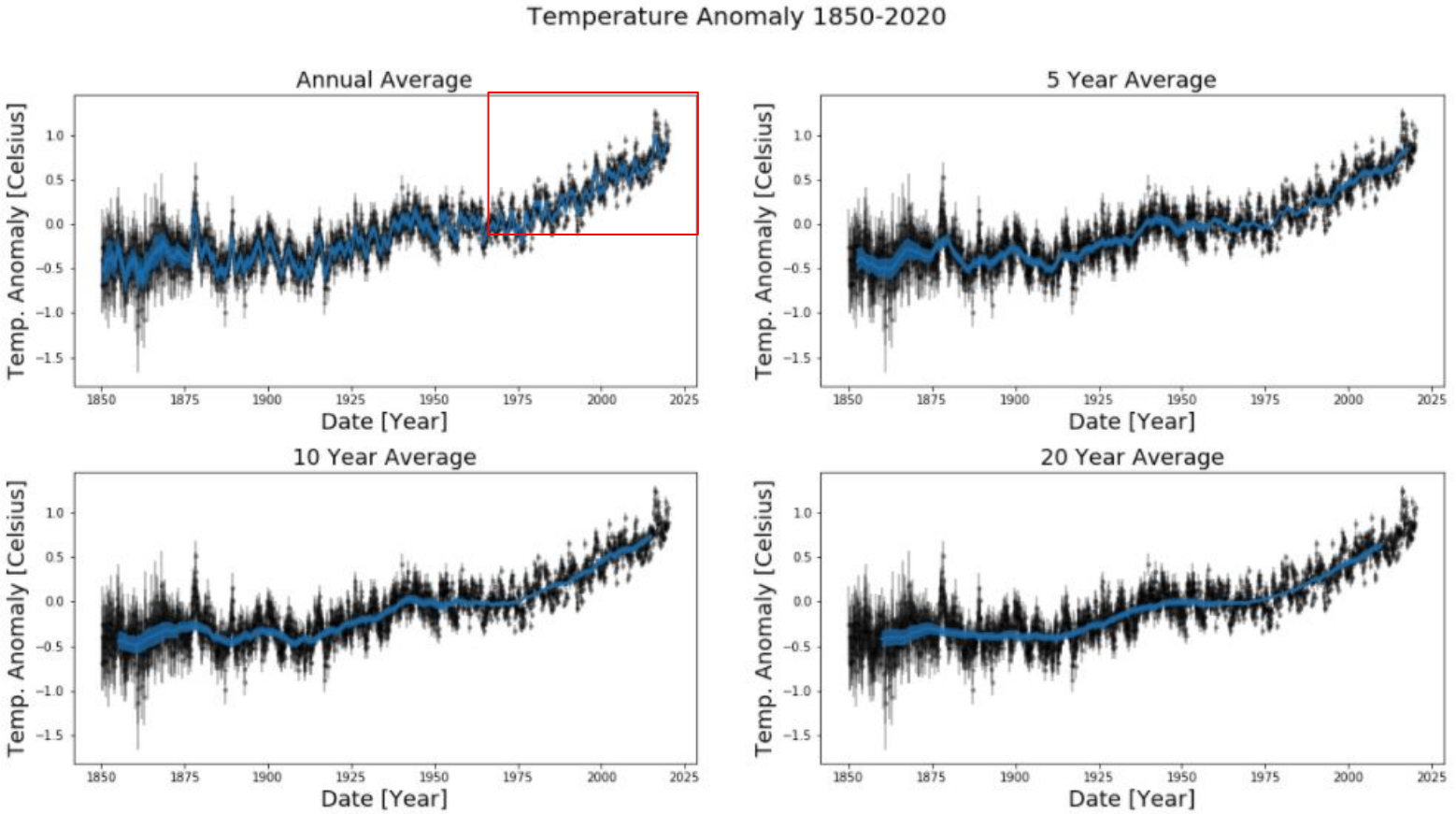




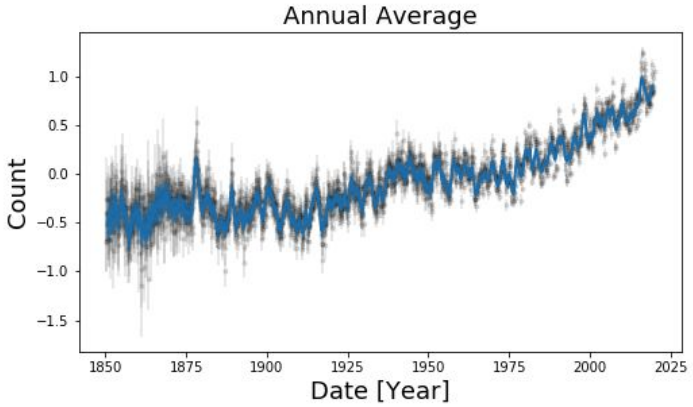
Connection to Global Warming

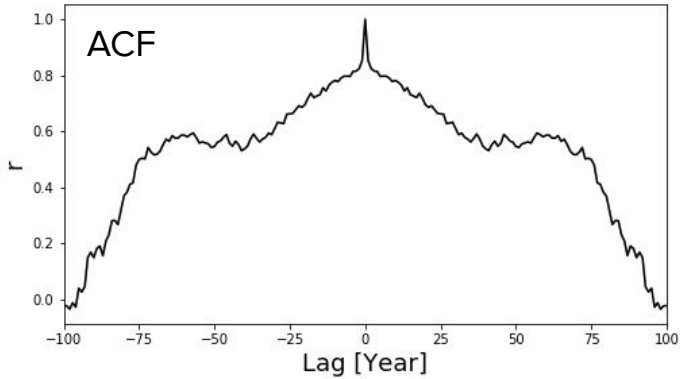
- Cross-correlation with earth surface temperature data
- GP modeling and forecasting for temperature data

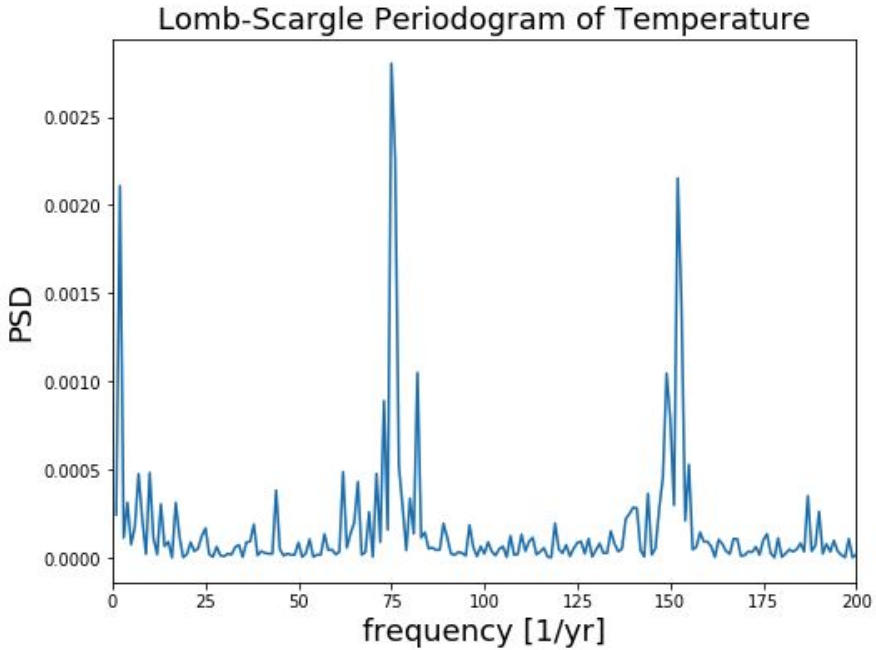
Again, LOOK AT THE DATA



Periodicity in Temperature

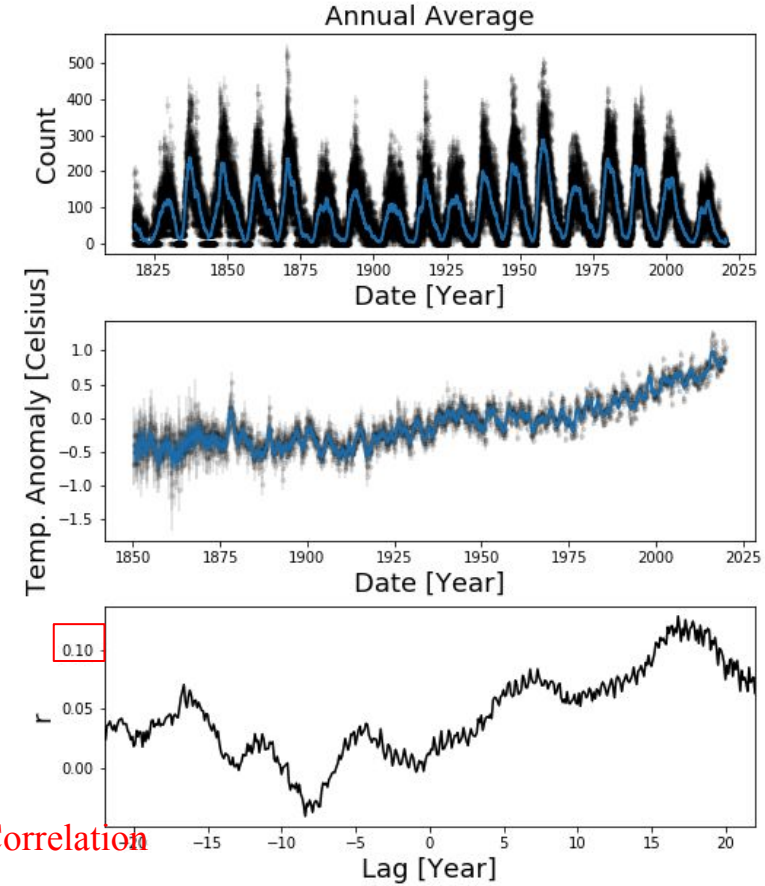


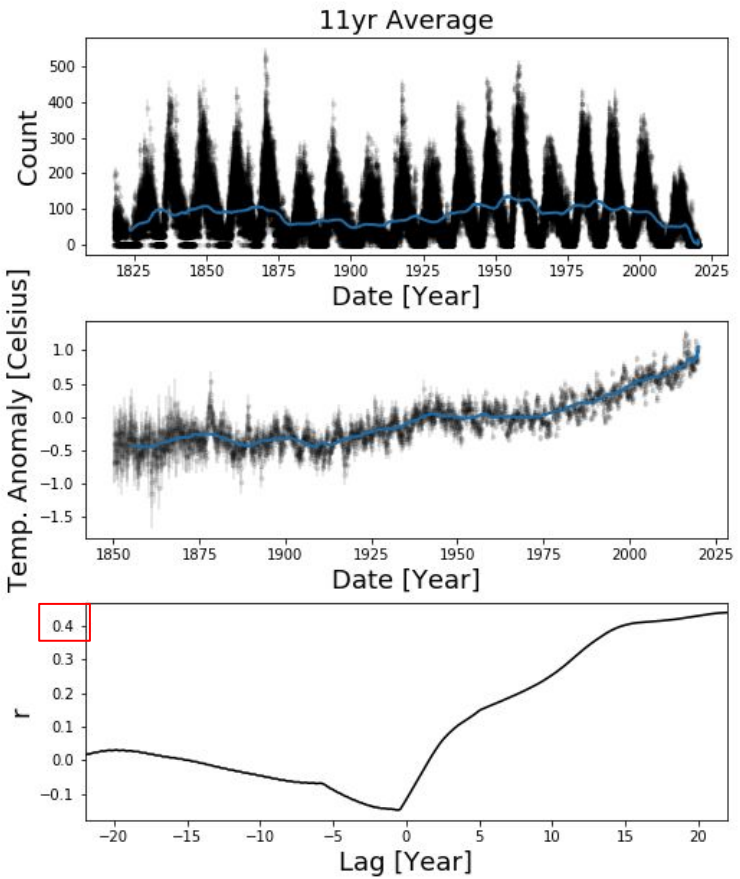




Short Term: ~1 yr
Long Term: ~75 yr

Any correlation between them? ➔ Cross-correlation!

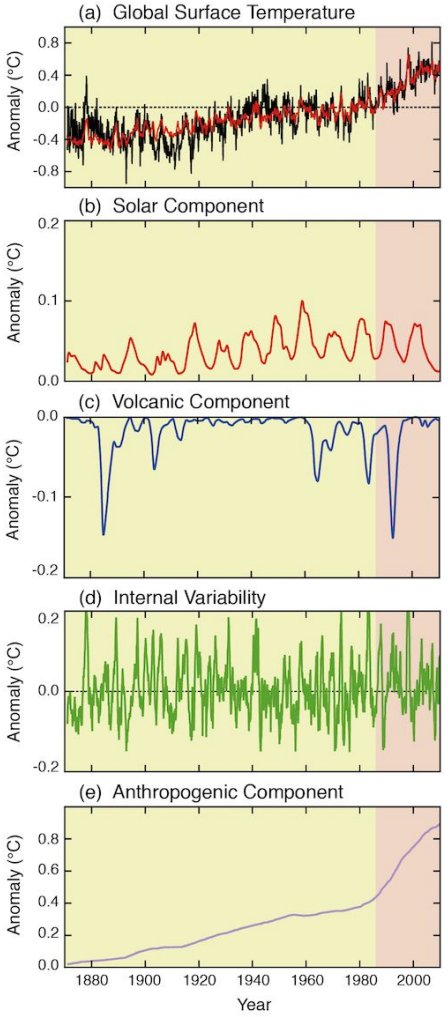


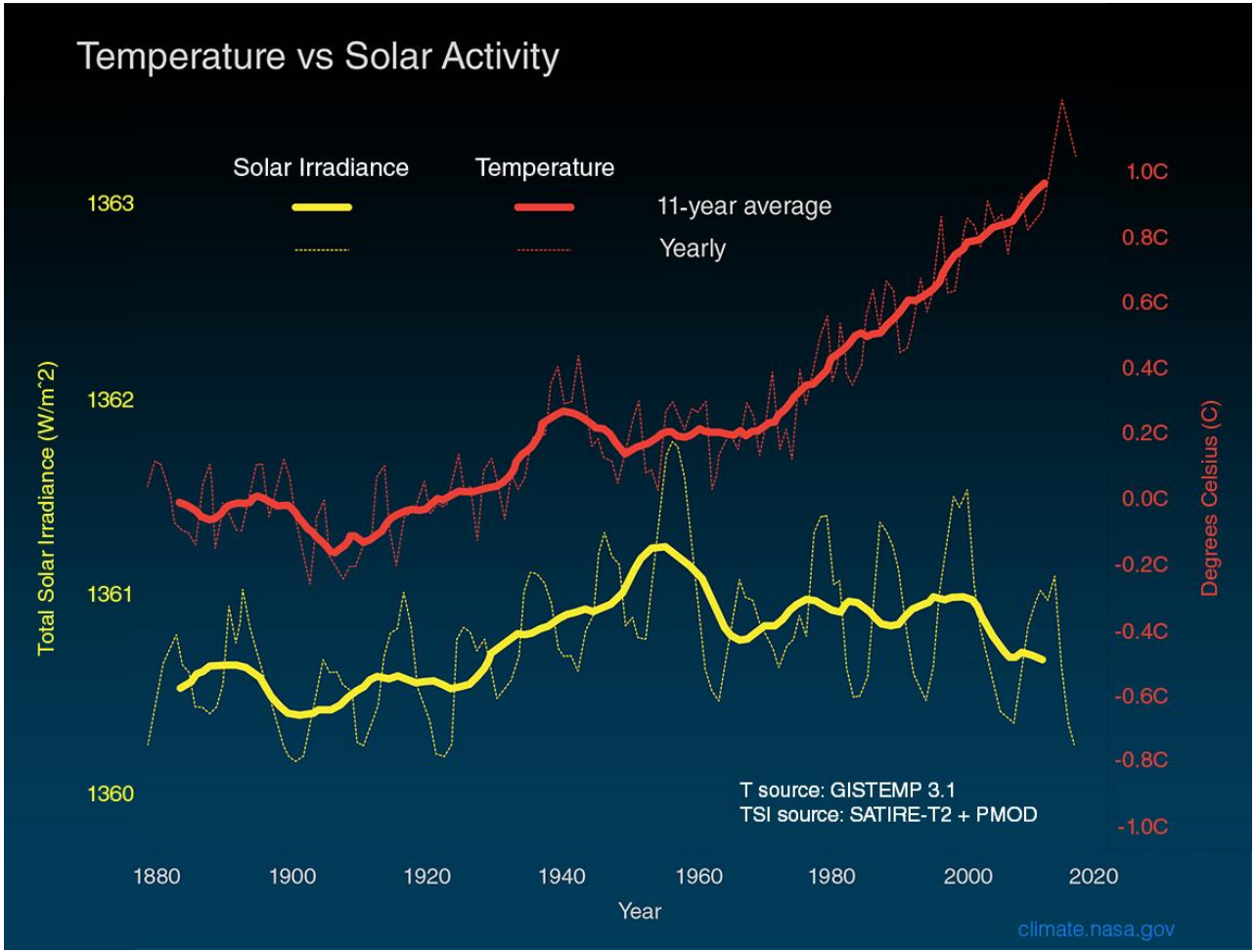


Weak Correlation

Why did this failed?

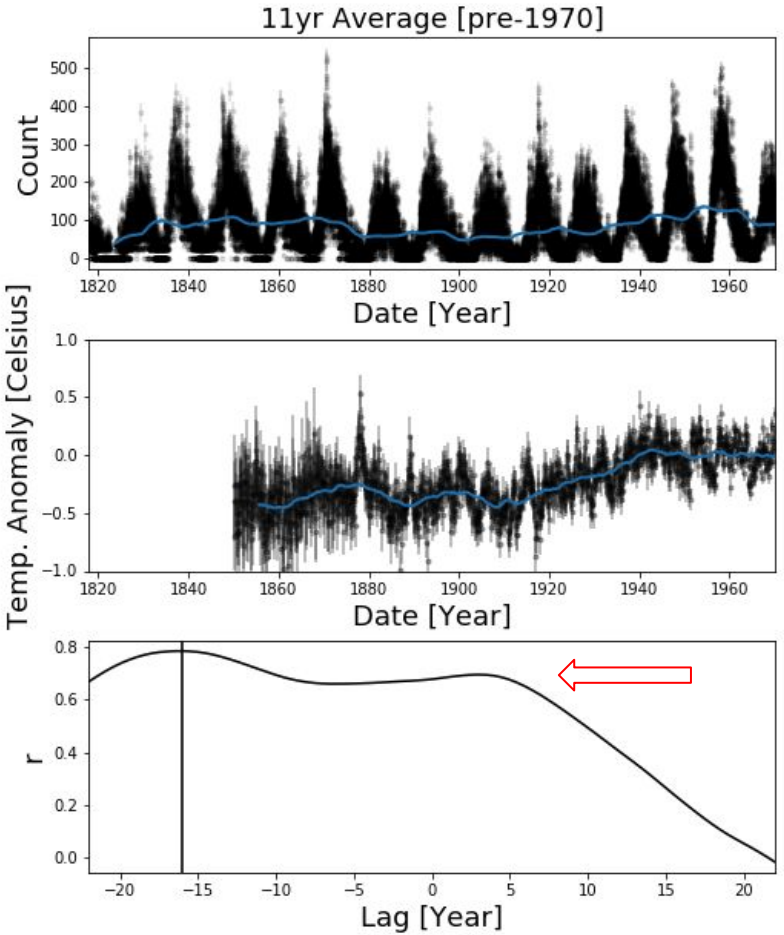
- Sunspot number is related to solar activity, but not a good/direct indicator of solar irradiance
- There are other factors in solar irradiance, e.g. Facula (bright spots), different wavelength have different effects on the Earth's atmosphere, some wavelengths may be out of phase with the solar cycle, … etc. [IPCC]
- Signal buried by more important factors (SSN contributes ~0.1 degC)



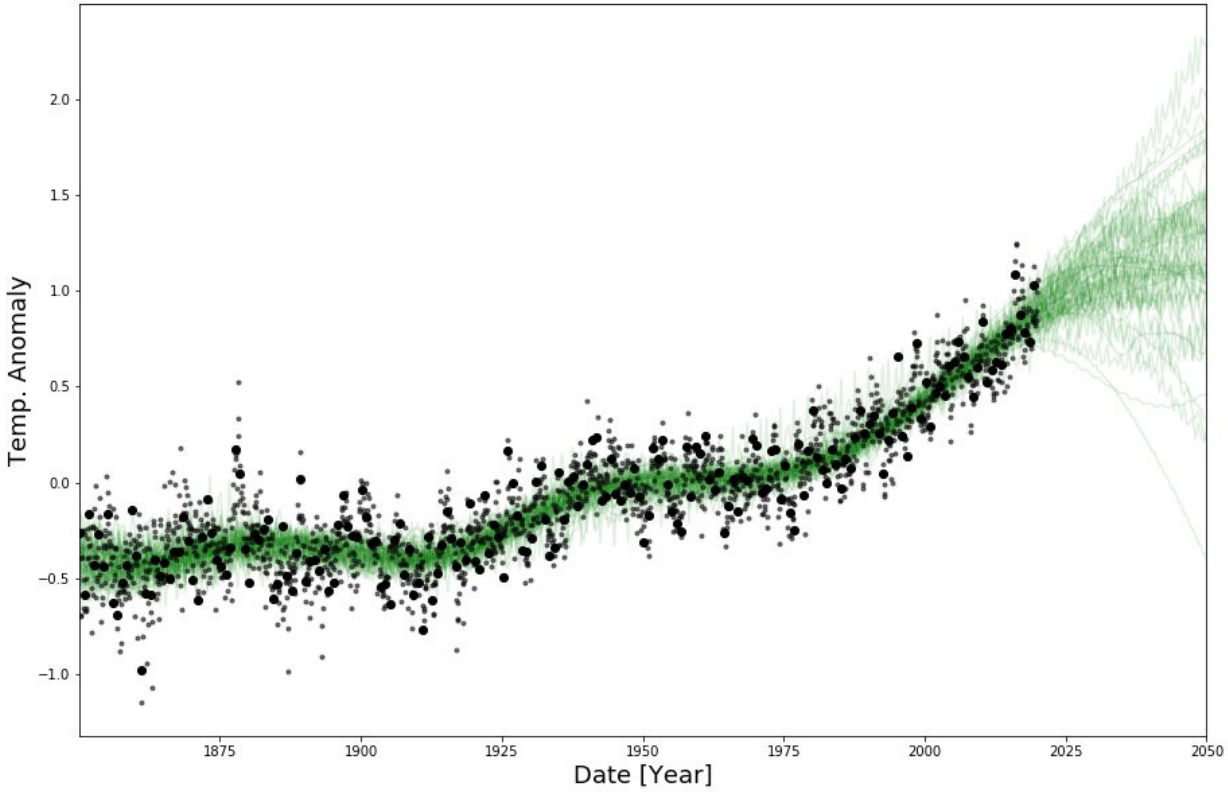

Temperature vs Solar Activity
Solar Irradiance
Temperature
11-year average
Yearly
1363
1362
1361
1360
Total Solar Irradiance (W/m^2)
1.0C
0.8C
0.6C
0.4C
0.2C
0.0C
-0.2C
-0.4C
-0.6C
-0.8C
-1.0C
Degrees Celsius (C)
T source: GISTEMP 3.1
TSI source: SATIRE-T2 + PMOD
1880
1900
1920
1940
1960
1980
2000
2020
Year
climate.nasa.gov

Pre-1970 data

Li et al. 2018 found a lag of 3 years between the long term SSN (22 year cycle) and temperature anomaly in pre-1970 era



Model temperature anomaly anyway...



Conclusion

This is actually really difficult

Modeling and Prediction of Sunspot Cycles

by

Li He

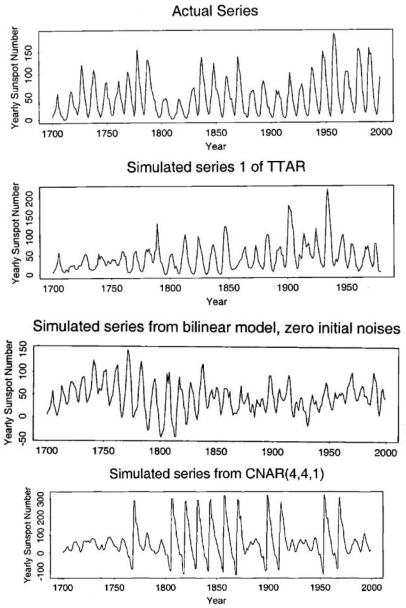
B.S., University of Science and Technology of China, June 1997

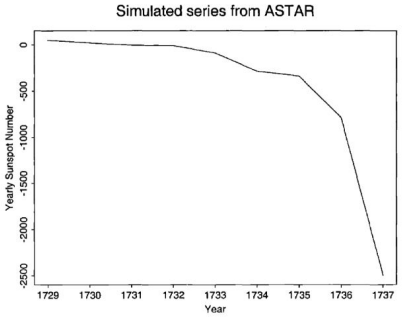
Submitted to the Department of Mathematics
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
at the
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
September 2001

© 2001 Li He. All rights reserved.
The author hereby grants to MIT permission to reproduce and to distribute publicly paper and electronic copies of this thesis document in whole or in part, and to grant others the right to do so.

11.4 Long-term simulations beyond one cycle

Given a few initial values, we simulated the models, trying to see whether the simulated series has the same shape as the real data. It turned out that for long term simulations, the simulated series do not capture the shape of the actual series. For long-term simulations no model of those we simulated really does well.





← lol