Dynamo Dynamics: On the Generation of Sunspots

Aidan Nakhleh and Jenna Gorham

Introduction

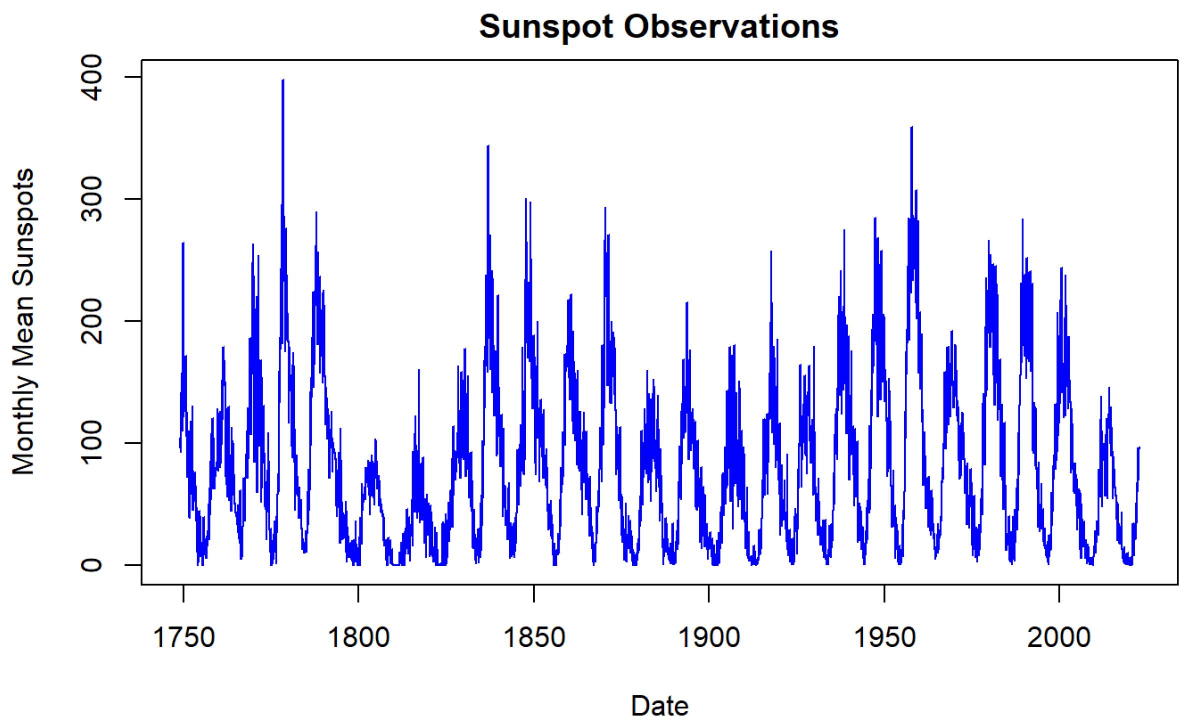
Sunspots are a byproduct of the solar cycle, a phenomenon that relates to magnetic fields produced from the Sun. Because the Sun is an electrically charged sphere of gas, rapidly moving charges in the Sun induce magnetic fields, visible from the surface of the Sun. Charged particles run along these magnetic field lines and when they recombine with each other to form a complete loop, massive amounts of energy are released, flinging particles off the surface of the Sun in processes known as solar flares and coronal mass ejections. These eruptions are often marked by sunspots, prominences on the Sun that appear darker than their surroundings because they are cooler. Because the Sun's magnetic field flips every 11 years on average, the solar cycle has a periodicity of 11 years as well. During times of maximal solar activity, there are more eruptions from these magnetic fields, and hence more sunspots. Observations of these sunspots are an easy way to track the solar cycle from Earth, and therefore this phenomenon has been monitored for hundreds of years. There are several databases keeping records of these observations to see this 11 year cyclic pattern over long periods of time. The solar cycle is known to be roughly 11 years, but how accurate is this trend over time? Has it stayed the same for hundreds of years? Are there cycles within this 11 year periodicity where the maximum count of sunspots varies with time? Does the invention of newer telescopes and means of detection have an effect on the measured periodicity? These are all questions that can be answered through the statistical analysis of these datasets.

Methods

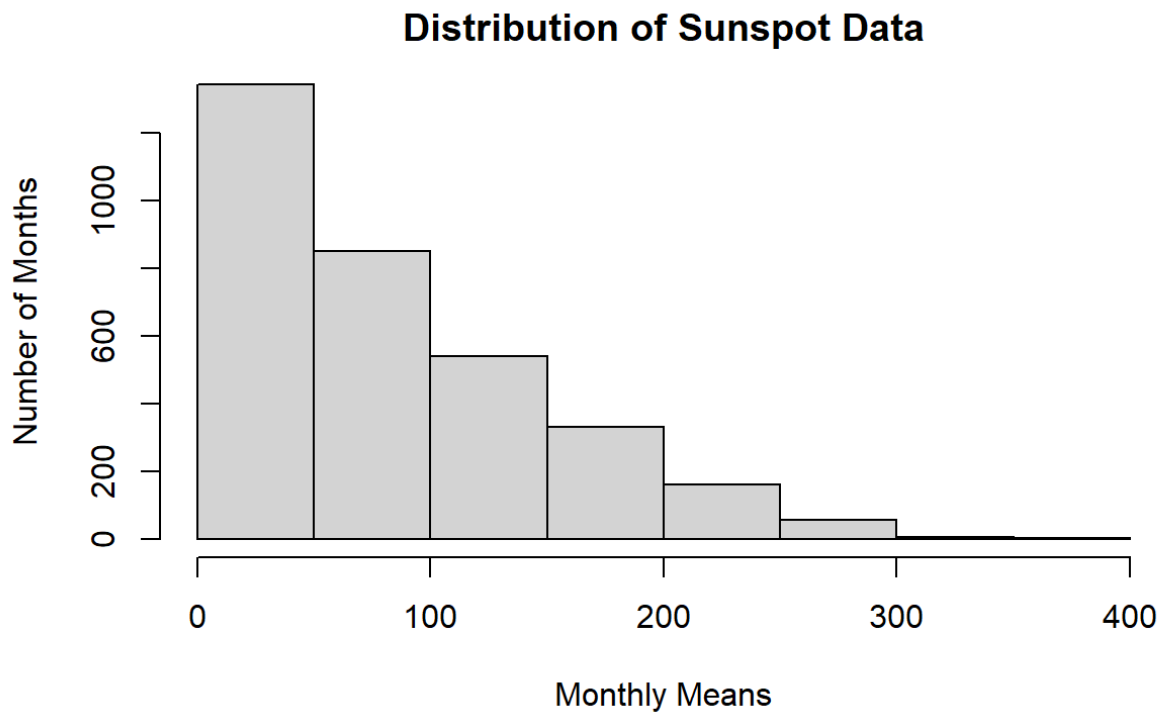
This analysis will be an observational study regarding the number of sunspots observed each month. Our dataset will be downloaded from the Solar Influences Data Analysis Center (SIDC) under the division of the Sunspot Index and Long Term Solar Observations (SILSO) division. The SIDC is the solar physics research department of the Royal Observatory of Belgium. This dataset is publicly available and can be downloaded from the SIDC website. For each month from January 1749 through September 2022, the datasets report the monthly mean total sunspot number and the number of observations used to compute this mean. The monthly mean total sunspot number is obtained by taking the sample mean of the daily total sunspot number of all days in the month for which observations were recorded. The primary source of possible error in the data is that observational technology has improved significantly from 1749 to modern day, so earlier data may not be fully accurate. We will utilize hypothesis testing to investigate trends within sunspot cycles.

Results

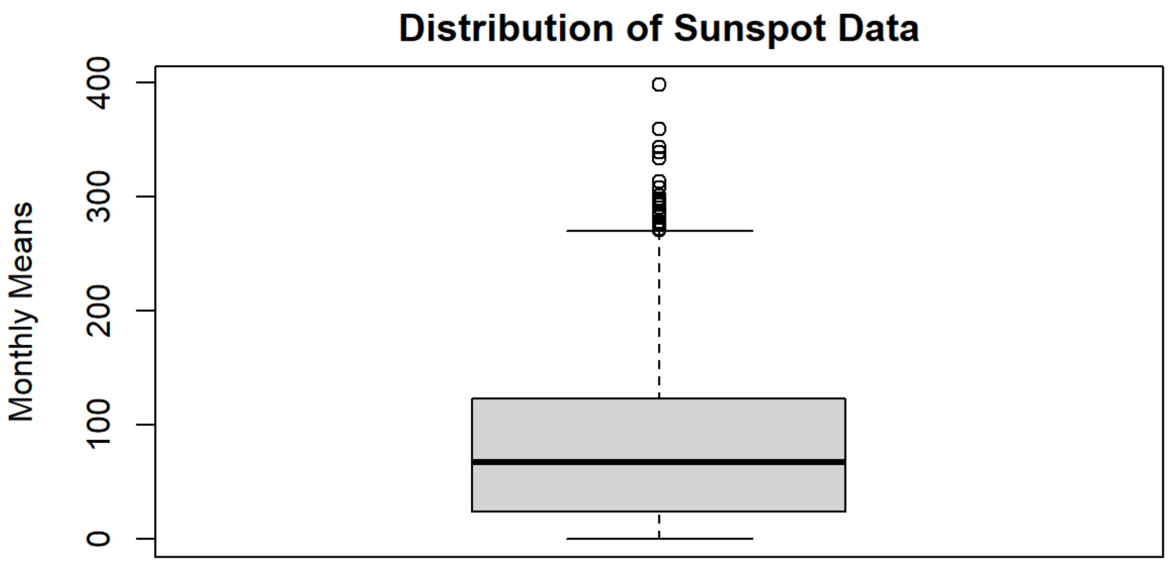
We begin with some simple visualizations to examine the trends within the sunspot data. Scientific consensus is that the occurrence of sunspots increases and decreases periodically according to an 11-year cycle. Graphing the monthly mean sunspot count as a function of time provides a clear representation of this periodicity. We collect this data from the Sunspot Index and Long-term Solar Observations organization, a World Data Center for the production, preservation and dissemination of the international sunspot number [1]. It is immediately evident that the number of sunspots periodically peaks at a local maximum and then decreases again. There also appears to be some variation in these cycles; for example, some cycles appear somewhat shorter or longer than others, and more notably, some cycles have much higher maxima than others. This investigation is centered around an examination of those differences.



In examining these differences, it is useful to gain a greater understanding of the spread of the data. One easy way to visualize this is with a histogram, shown below. This histogram illustrates that the vast majority of months have a relatively low number of sunspots. The histogram has a peak at 0 to 50 sunspots with fewer and fewer months in which a larger number of sunspots were observed. The distribution is significantly skewed to the right, indicating again that very few months had a large number of observed sunspots.

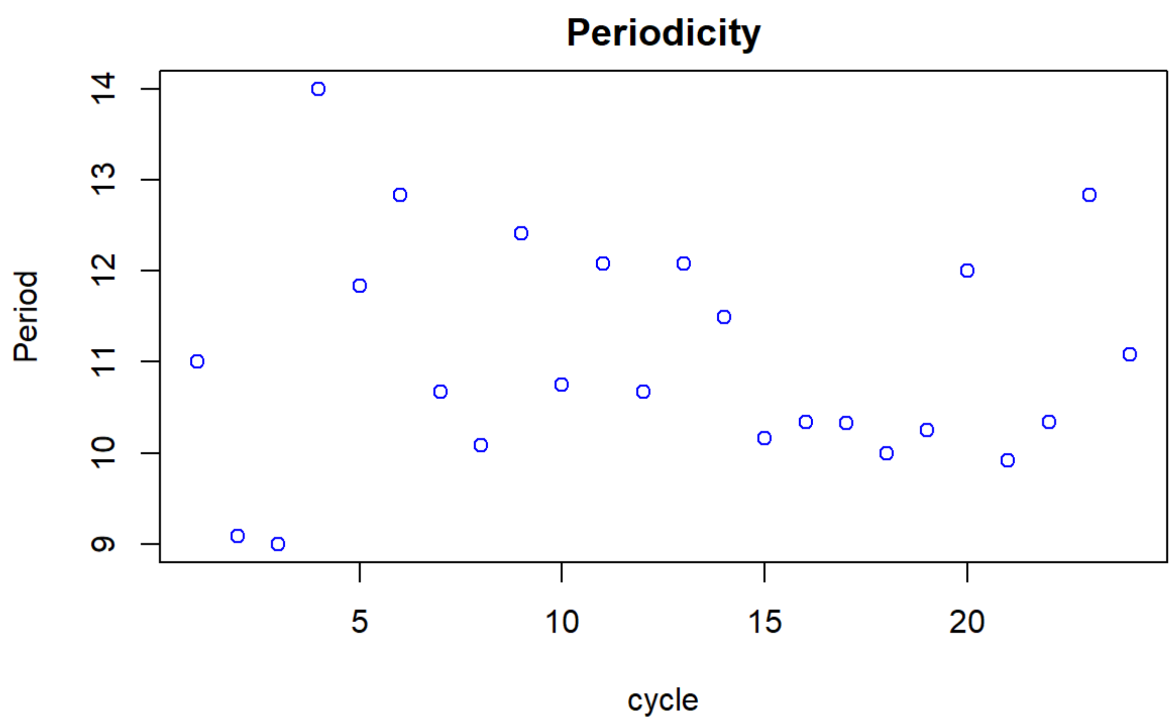


We can also visualize the spread of the data through a boxplot. Note that the median number of observed sunspots is relatively small, indicating that the vast majority of months had relatively few sunspots. Precisely, the median is 67.05 sunspots, indicating that fewer than 67.05 sunspots were observed in half of the months. This stands in stark contrast to the maximum, 398.2 sunspots, which were observed in May 1778. We can see that there are several outliers in the upper end of the distribution. These points correspond to cycles in which there were greatly more sunspots than usual observed at the maximum of the distribution.



From the data of monthly mean sunspot observations, we determine the following: for each cycle of sunspot observations, we determine both the maximum value of the cycle and the length of the sunspot cycle, calculated by finding the length of time between successive local minima. Then we proceed to analyze these data as a function of time and examine their spread.

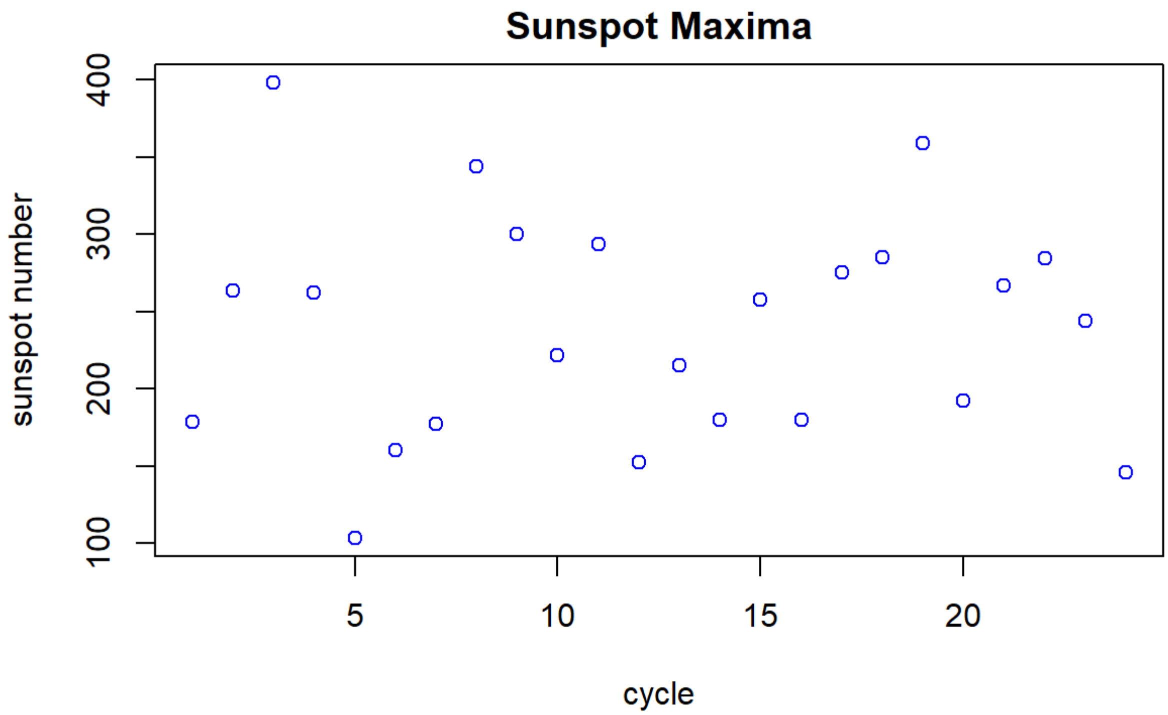
By analyzing the timestamp of the minima in each cycle and subtracting the start minimum time from the end minimum time we can get an estimate for the period of each cycle. Doing so and plotting the period vs cycle number, we see the following:



This plot demonstrates how the period of each sunspot cycle changes for each cycle. Cycle 1 starts in 1755 and cycle 24 ends in 2020. Consistent observation of sunspots in this dataset began in 1749 but we decided to start our analysis at the beginning of the first full cycle with a minimum, and end at the end of the minimum of the last full cycle in 2020.

The mean of these periods is 11.05 years and the data range from just over 9.002 years to 13.998 years. This agrees with NASA reports and another analysis performed by David Hathaway et. al. [2]. Our standard deviation is also quite large at 1.242 years. This implies that there is some large variation, so further analysis will likely be necessary. Hathaway et. al. report several attempts to fit how the sunspot number rises and falls with time, and the system does appear to be highly nonlinear which can explain the large standard deviation.

We can also plot the sunspot maximum value for each cycle:



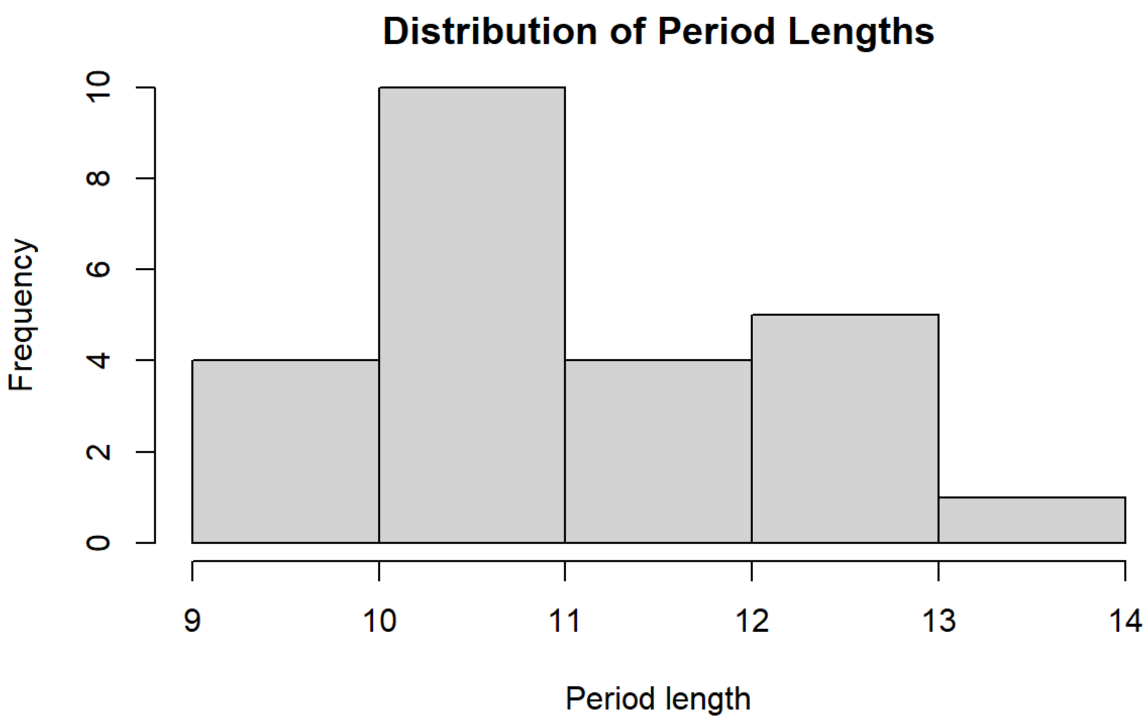
The mean maximum value is 239.3375 sunspots, and the standard deviation is 73.24972 which is also very large.

Analysis

We wish to test whether the true mean length of the solar cycle, as indicated by sunspot observations, is indeed 11 years as the literature suggests. The appropriate test to perform in this context is a two-sided one-sample t-test. We perform a t-test because the true standard deviation in the length of the solar cycle is not known.

In order for a t test to apply, we need to first verify that the distribution of cycle lengths is approximately normal. Otherwise, the central limit theorem does not apply and a t procedure is not appropriate.

Our data is based on a sample of 24 observed complete solar cycles. Below is a histogram showing the distribution of the data for cycle lengths.



Note that the distribution is relatively symmetric and approximately normal. Since there are no outliers or strong skewness in the data, given a sample size of 24 observations, a t procedure is appropriate in this instance. We might expect that with more observations, according to the law of large numbers, this distribution would look even closer to a normal distribution. However, our data is limited because the length of a solar cycle is so long and within this context, the hundreds of years over which data have been collected is a relatively short timespan.

With the necessary setup out of the way, we can proceed with the formal hypothesis test. Our null hypothesis is that the true mean length of the solar cycle is 11 years as the scientific literature indicates. The alternative hypothesis is that the mean length is not 11 years. After calculating the t-test statistic, we obtain a p-value of 0.8389. Accordingly, at the 5% significance level, we fail to reject the null hypothesis. Since the p-value is large, there is significantly strong evidence that the null hypothesis is accurate in its description of the true state of nature; that is, there is strong evidence that the mean length of the solar cycle is indeed 11 years. We also calculate the 95% confidence interval to be (10.52766, 11.57659). That is, we have 95% confidence that the true mean length of the solar cycle is between about 10.53 and 11.58 years. We note that this is a relatively narrow range of about one year, and again conclude that we are highly confident that the true mean length of the solar cycle is close to 11 years.

Conclusion

In this analysis, we wished to test the accuracy of sunspot fluctuations over time in relation to the global solar cycle. The solar cycle is dependent on the number and strength of magnetic fields in and around the Sun. As the solar cycle progresses, magnetic fields on the Sun's surface fling material off, resulting in granulations of cooler temperatures which appear darker. These are sunspots. When the solar cycle reaches its maximum, the number and strength of magnetic fields increases, so the number of observable sunspots also increases. When the solar cycle reaches its minimum, the number of sunspots also decreases. Therefore, we tested to see if the length of the solar cycle is the same as the period of a sunspot cycle, from minimum sunspot number to the next minimum number. NASA estimates that the solar cycle on average lasts 11 years, so the null hypothesis is that the mean period of sunspots that have been observed since the mid 18th century have this same period. By analyzing this historical data, we have found that the mean period of sunspots is 11.05 years, and the sample standard deviation is 1.242 years. We created a histogram of these periods to see their peak frequencies, and deemed that the distribution was symmetric enough to run a statistical analysis. Because we do not know the population standard deviation, we will use a t test to determine if we reject or fail to reject the null hypothesis. Doing so, we found that the probability (p-value) of observing a mean of 11.05 years when the null hypothesis mean is 11 years is 0.8389. For a result to be statistically significant, the p-value must be less than 0.05, so we fail to reject the null hypothesis. This is not unreasonable, seeing as sunspots have been a method to estimate the solar cycle for a very long time. Modern measurements done by space agencies utilize solar probes and advanced magnetometers to measure magnetic field strengths on the Sun directly, along with numerical simulations, and they still find the solar cycle to be around 11 years. Failing to reject the null hypothesis of 11 years indicates that sunspots are still a good metric to study the solar cycle, and historical predictions have withstood centuries of improvement. Sunspot data should therefore continue to be collected and analyzed, so that standard deviations can be reduced and more accurate gauges of the solar cycle can be determined.

References:

Sunspot Index and Long-term Solar Observations. “Monthly Mean Total Sunspot Number.” *World Data Center for the Production, Preservation and Dissemination of the International Sunspot Number*, SILSO, https://www.sidc.be/silso/infosnmtot.

Hathaway, David H., Robert M. Wilson, and Edwin J. Reichmann. "The shape of the sunspot cycle." *Solar Physics* 151.1 (1994): 177-190.