Monthly mean total sunspot number analysis

Time Series and Forecasting - Final Project

Klismam F. Pereira

23/12/2020

Introduction

Sunspots are earth sized dark areas that occur at the surface of the Sun, and are commonly found in pairs that have magnetic fields pointing in diverging directions. In them, the magnetic field is higher than anywhere else on the Sun's photosphere (i.e. surface), and 2500 times stronger than Earth's own field. They appear as darker spots because they're colder than the surrounding area - while sunspots temperature is around 3593.33 ºC, the Sun's surface is around 5537.78 ºC. These spots are usually compose of a darker region called *umbra*, and a lighter surrounding region called *penumbra* (Weather Forecast Office, National Weather Service n.d. [1]).

Sunspots are strongly related to **solar flares** and to **geomagnetic storms** that occur here on earth. Solar flares are gigantic explosions that happen near sunspots, and can release up to a billion megatons of TNT of energy. They also emit x-rays and magnetic fields that cause said geomagnetic storms. As a result, during sunspot maxima, more solar flares and more geomagnetic storms happen. It is quite important to understand when said events take place, since **their effects on our daily lives can be quite disruptive**, among which are the increase in Northern and Southern lights, possible disruption in radio transmissions and power grids, damaging electronics in satellites, etc (Weather Forecast Office, National Weather Service n.d. [1]).

Given their impact, it is needless to stress how important it is to understand their mechanisms. For that end, the **sunspot number** is used. The sunspot number, also called sunspot index, is the oldest solar activity index used to characterize the Sun's **eleven year cyclical behavior**. It can be found in the abstract of several papers, and has application in many fields of study such as climatology, meteorology, space physics, etc (Weather Forecast Office, National Weather Service n.d. [1], Berghmans et al. n.d. [2]).

It's history begins when sunspots started being systematically observed in the 17th century after the invention of the telescope. Years later, in 1843 Heinrich Schwabe published his discovery on the ten-year cycle of the number of sunspots, as a result of studying his recordings of the number of sunspots observed daily since 1826. The effects of the solar magnetic cycle had already been observed in 1803, though, when Ritter reported that auroras are more commonly sighted during specific time intervals, nowadays called solar maxima (Berghmans et al. n.d. [2]).

Only in 1852 Ewdward Sabine, chief British promoter of magnetic studies, pointed out the "coincidence" between the period and epochs of minima and maxima described by H. Schwabe and the ones found for magnetic variation, and in the same year Rudolf Wolf presented his results showing that the actual period was of 11.1 years. He also derived spot numbers from scattered data back to 1749. Later, with more observatories at disposal and interested in the matter, the continuity of counts is preserved through joint effort (Berghmans et al. n.d. [2]).

In 1980, the production of the sunspot index, i.e. the **International Sunspot Number** becomes responsibility of the Sunspot Index Data Center (SIDC), of the Royal Observatory of Belgium. Under their administration, the number of contributing stations doubled in a few years, improving accuracy and stability of the international sunspot number, and in the 2000, the SIDC became a Regional Warning Center for the Western Europe, which is a space weather forecast and monitoring center of the International Space Environment Service (Berghmans et al. n.d. [2]).

Space weather is an emerging science with great and direct relevance to technological systems. Beyond the already noted impacts, solar influences might have an important influence on the evolution of the earth's climate, thus understanding the solar cyclic activity is important to humanity as whole (Weather Forecast Office, National Weather Service n.d. [1], Berghmans et al. n.d. [2]).

Objective

Given the importance of such an interesting subject, in this project the aim is to explore the monthly mean total sunspot number using time series analysis tools, explain some of the series behavior by fitting basic models to it, and finally try to perform simple predictions based on the best of these models.

The time series

The time series consists of the monthly mean total sunspot number from 1/1749 to 11/2020. As described in the introduction section, the earliest records were compiled by Rudolph Wolf. The monthly mean total sunspot number is obtained by calculating the average of the number of daily total sunspot number over each month (Berghmans et al. n.d. [2], Sunspot data from the World Data Center SILSO [3]).

The data set contains seven features - Year, Month, Date in fraction of year, Monthly mean total sunspot number, Monthly mean standard deviation of the input sunspot numbers, Number of observations used to compute the monthly mean total sunspot number, Definitive/Provisional marker (1 means the value is definitive, 0 means its still provisional) (Sunspot data from the World Data Center SILSO [3]).

From the seven features, the only ones that are used in this project are:

- Year
- Month
- Monthly mean total sunspot number

Exploratory Data Analisys

In this section, the time series is explored graphically, and its main characteristics are described. The code chunk below loads the necessary libraries, and the next imports the data set and filters the features described in the previous section, it also creates a time series R object which comprises the information in the monthly mean total sunspot, Year and Month features. The monthly mean total sunspot number for the two first and last years are shown.

```r
library(astsa)
library(forecast)
```

```r
df <- read.csv2("SN_m_tot_V2.0.csv", header = FALSE,
                col.names = c("Year",
                              "Month",
                              "Date.fraction",
                              "mean.total.sunspot",
                              "mean.standard.deviation",
                              "Number.of.observations",
                              "definitive.marker"),
                colClasses = c("character",
                               "integer",
                               "NULL",
                               "character",
                               "NULL",
                               "NULL",
                               "NULL"))

df$mean.total.sunspot <- as.numeric(df$mean.total.sunspot)

first_year <- df$Year[1]
first_month <- df$Month[1]

sunspot.ts <- ts(df$mean.total.sunspot, start=first_year, frequency=12)

head(sunspot.ts, 24)
```

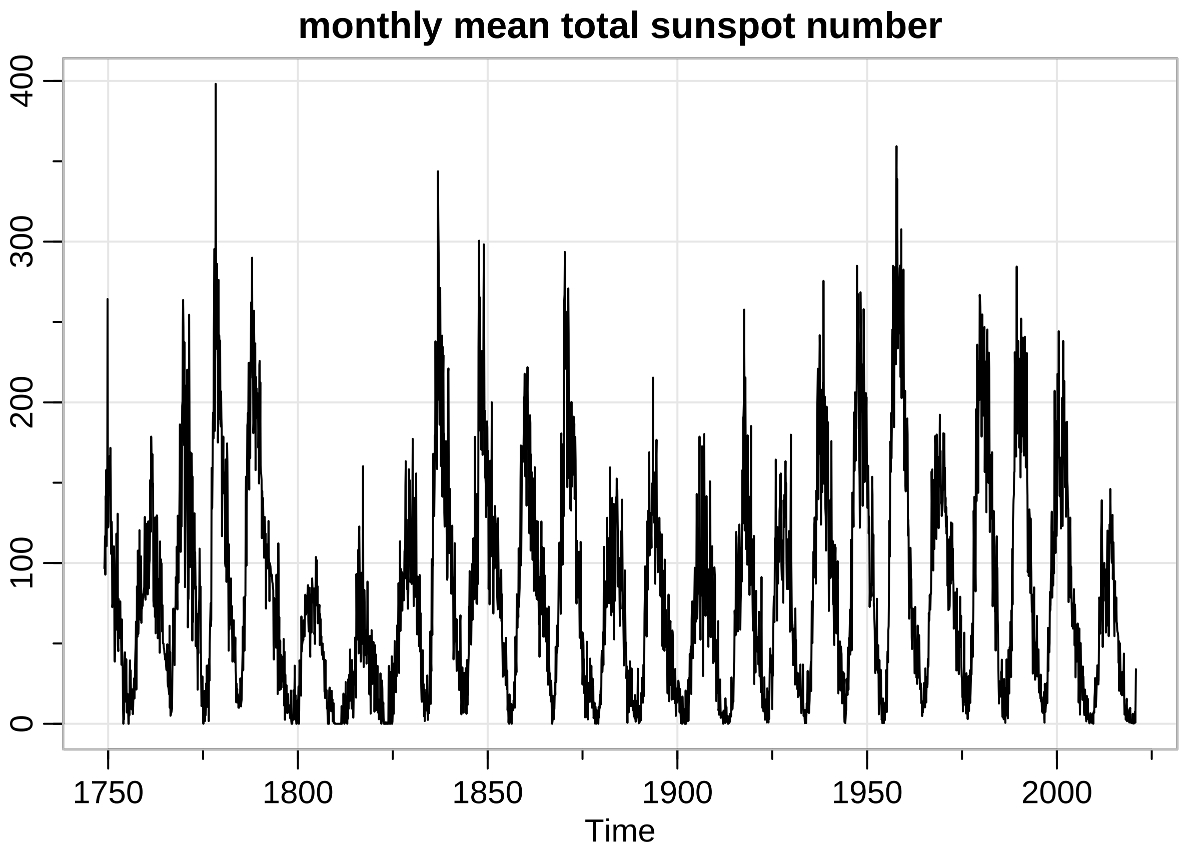
```
##        Jan   Feb   Mar   Apr   May   Jun   Jul   Aug   Sep   Oct   Nov   Dec
## 1749  96.7 104.3 116.7  92.8 141.7 139.2 158.0 110.5 126.5 125.8 264.3 142.0
## 1750 122.2 126.5 148.7 147.2 150.0 166.7 142.3 171.7 152.0 109.5 105.5 125.7
```

```r
tail(sunspot.ts, 23)
```

```
##       Jan  Feb  Mar  Apr  May  Jun  Jul  Aug  Sep  Oct  Nov  Dec
## 2019  7.7  0.8  9.4  9.1  9.9  1.2  0.9  0.5  1.1  0.4  0.5  1.5
## 2020  6.2  0.2  1.5  5.2  0.2  5.8  6.3  7.6  0.7 14.4 34.0
```

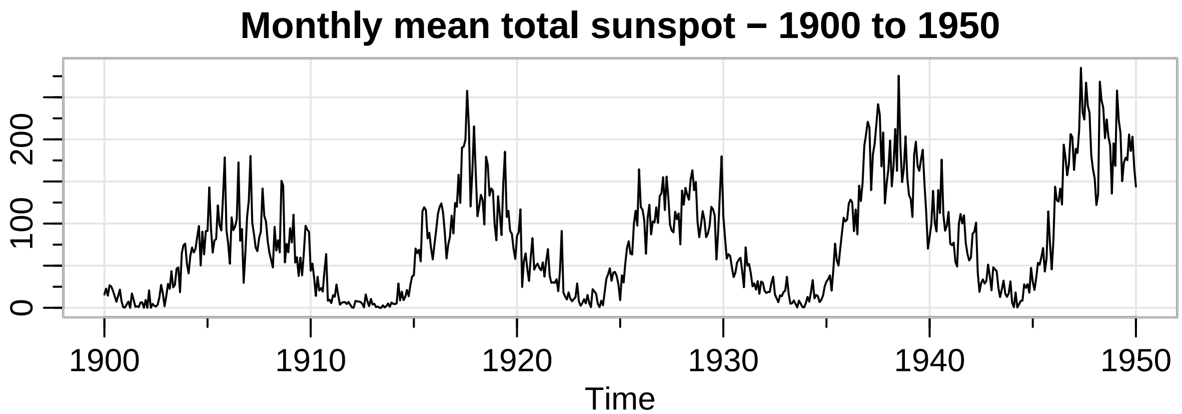
The following plot represents the complete time series. It seems 25 sunspot maxima occurred since 1750, and latest entries suggest the latest minimum is already past. The cycles observed are compatible with the long 11 year cycles description given in the introduction section. There is no clear trend, and the cycles aren't ruled by the more commonly found in time series yearly seasonality. Differences in the maxima magnitudes imply heteroscedasticity, or non-constant variance. Finally, there is no discontinuities or abrupt changes in this specific time period.

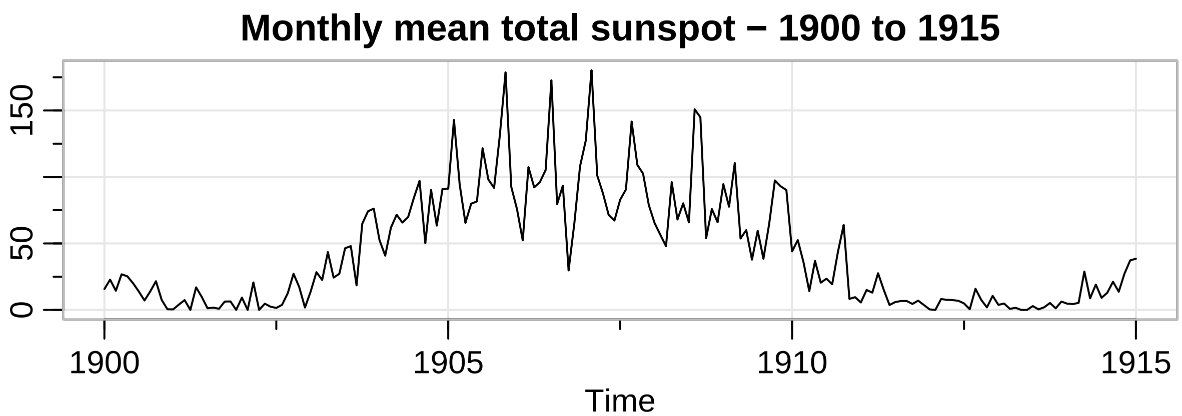
```r
tsplot(sunspot.ts, main="monthly mean total sunspot number", ylab = "")
```



Two smaller periods of the time series are plotted below. The first graphic presents a window from 1900 to 1950, where the 11 years cycle is more evident. The second focus on the period between 1900 and 1915, in what it seems to be a full cycle. There appears to be some cyclic behavior within the bigger 11 year cycle, but at this time it can't be pinpointed to yearly seasonality or other influences.

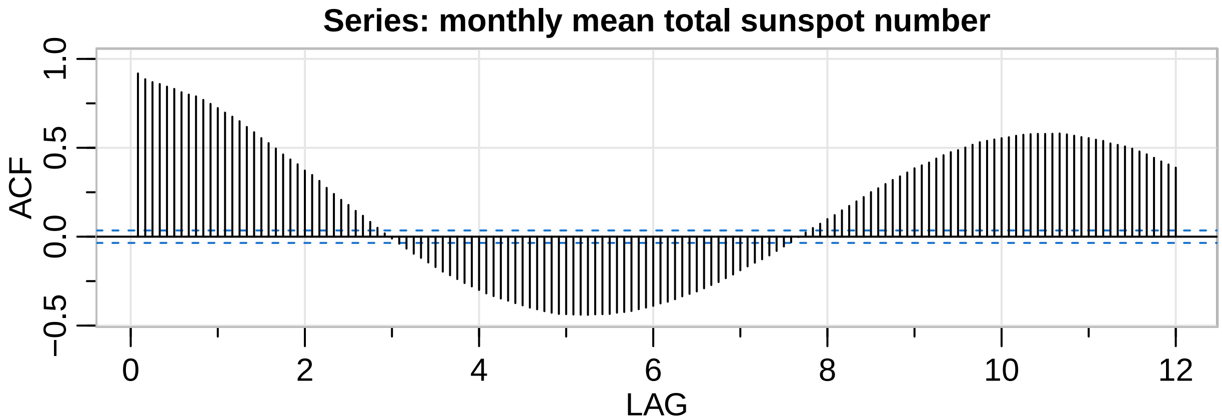
```
par(mfrow=c(2,1))
tsplot(window(x=sunspot.ts, start=1900, end=1950),
       main="Monthly mean total sunspot - 1900 to 1950", ylab = "")

tsplot(window(x=sunspot.ts, start=1900, end=1915),
       main="Monthly mean total sunspot - 1900 to 1915", ylab = "")
```

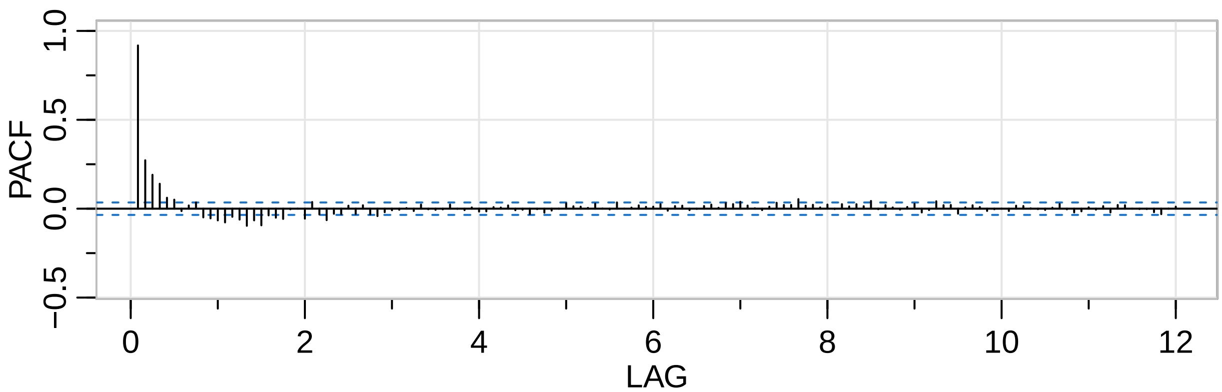




Now, both ACF and PACF graphics for 144 lags are shown (i.e. twelve years). They display a non stationary series, which was clear from the time series plot.

```
sunspot.acf.pacf <- acf2(sunspot.ts, max.lag=144,
                         main="Series: monthly mean total sunspot number")
```





The ACF presents high correlations that do not seem to be decaying to zero, even though 144 lags are shown. They change signals at around the three years mark (i.e. 36 lags) and reach a new negative maximum at around 5 years (i.e. 60 lags). The correlations change signal again and arrive at positive maximum close to 11 years (i.e. 132 lags). At a first glance, this ACF plot may imply some of the long cyclical behavior described in previous paragraphs, with high correlations occurring 5 and 11 years apart.

The PACF presents high correlations at lags one to four, and there are also smaller partial correlations that are outside the significance threshold for stationarity in the negative side of the plot.

Seeing that the time series correlations in the ACF plot are not decaying to zero, and that it is not clear from the PACF plot which degree of a simple auto-regressive (AR) model would be an interesting first approach, some more exploratory data analysis and processing are required, which are the next steps of this project.

REFERENCES

[1] Weather Forecast Office, National Weather Service n.d., *The Sun and Sunspots, Sioux Falls, SD, United States of America*, viewed 23 December 2020, https://www.weather.gov/fsd/sunspots.

[2] Berghmans, D.,Van der Linden, R.A.M., Vanlommel, P., Clette, F., Robbrecht, E. n.d., *History of the Sunspot Index: 25 years SIDC*, SIDC, Royal Observatory of Belgium, Ringlaan -3- Av. Circulaire, B1180 Brussels, Belgium, viewed 23 December 2020, http://www.sidc.be/silso/IMAGES/about/Berghmansetal2006.pdf.

[3] *Sunspot data from the World Data Center SILSO*s, Royal Observatory of Belgium, Brussels, viewed 23 December 2020, http://www.sidc.be/silso/datafiles.