Skolkovo Institute of Science and Technology

Experimental Data Processing
MA060238

The relationship between solar radio flux F10.7 and sunspot number

Authors:

Luca Breggion
Irina Yareshko
Ivan Kozhin

October 10, 2022

1 Introduction

The objective of this assignment is to understand the relationship between main indicators of solar activity, sunspot number and the solar radio flux at 10.7 cm (2800 MHz) by applying multi-dimensional linear regression technique.

2 Plots of the monthly mean sunspot number and solar radio flux F10.7

The data provided contain information about:

- year of measurements
- month of measurements
- monthly mean solar radio flux at 10.7 cm
- monthly mean sunspot number

In accordance with the design requirements, a plot with original data was built. The result is shown in **Figure 1**.

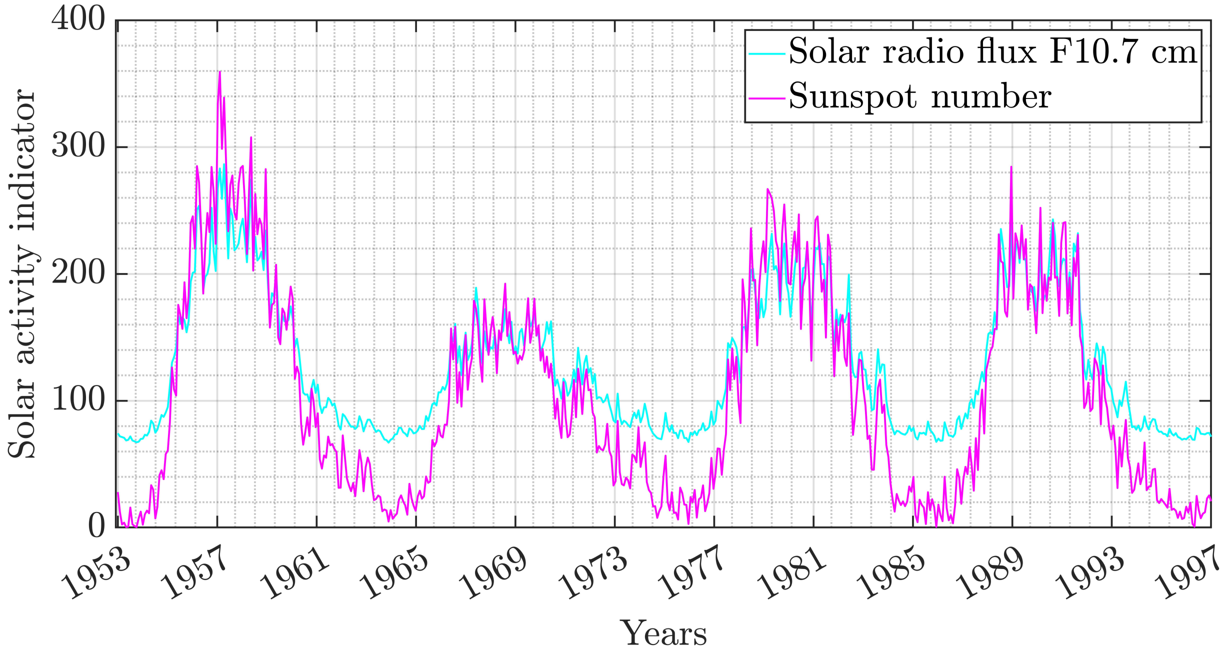

Figure 1: Monthly mean sunspot number and solar radio flux F10.7

These data contain a large oscillatory component caused by the presence of noise during measurements. Therefore the data need to be processed before any analysis.

3 Scatter plot between monthly mean sunspot number and solar radio flux F10.7

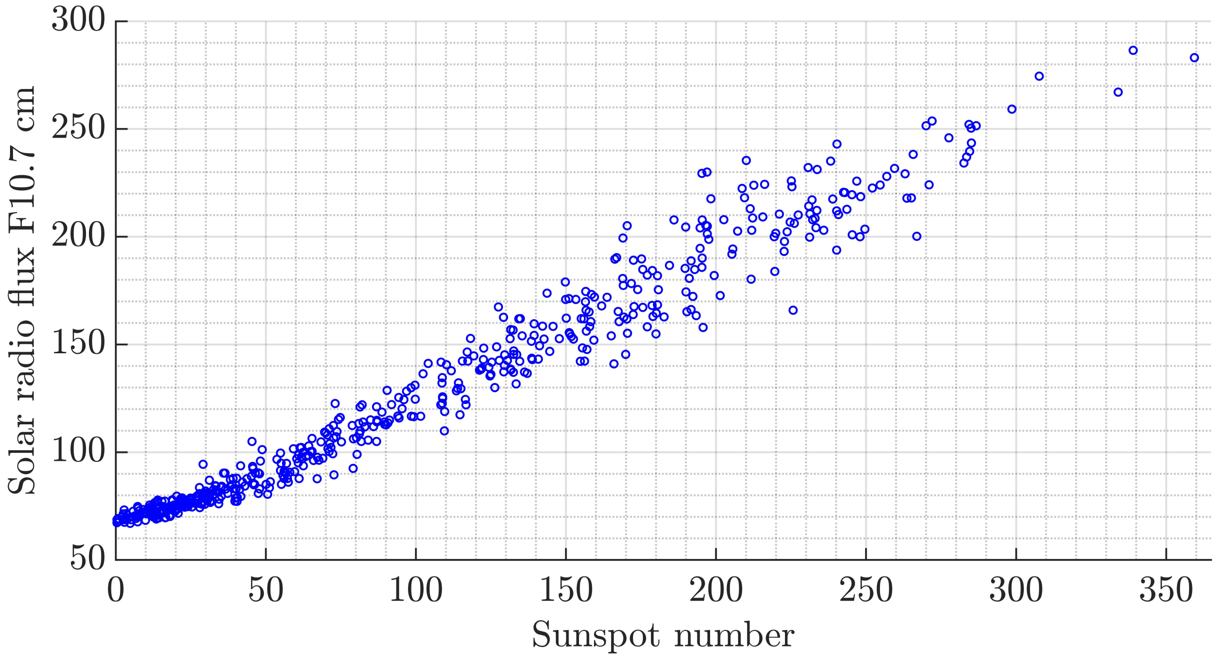

Figure 2: Scatter plot between monthly mean sunspot number and solar radio flux F10

It was observed from the scatter plot that there is a strong linear relationship between monthly mean sunspot number and solar radio flux F10.7, namely with the increase of the sunspot number, the solar radio flux increments in a linear way.

4 13-month running mean smoothing

In order to smooth the data, the 13-month running mean method was used with the following formula:

$$\mathbf{R} = \frac{1}{24}R_{i-6} + \frac{1}{12}(R_{i-5} + R_{i-4}... + R_{i-1} + R_i + R_{i+1} + ... + R_{i+5}) + \frac{1}{24}R_{i+6} \tag{1}$$

After the calculations, a new figure is obtained. The result is shown in **Figure 3**.

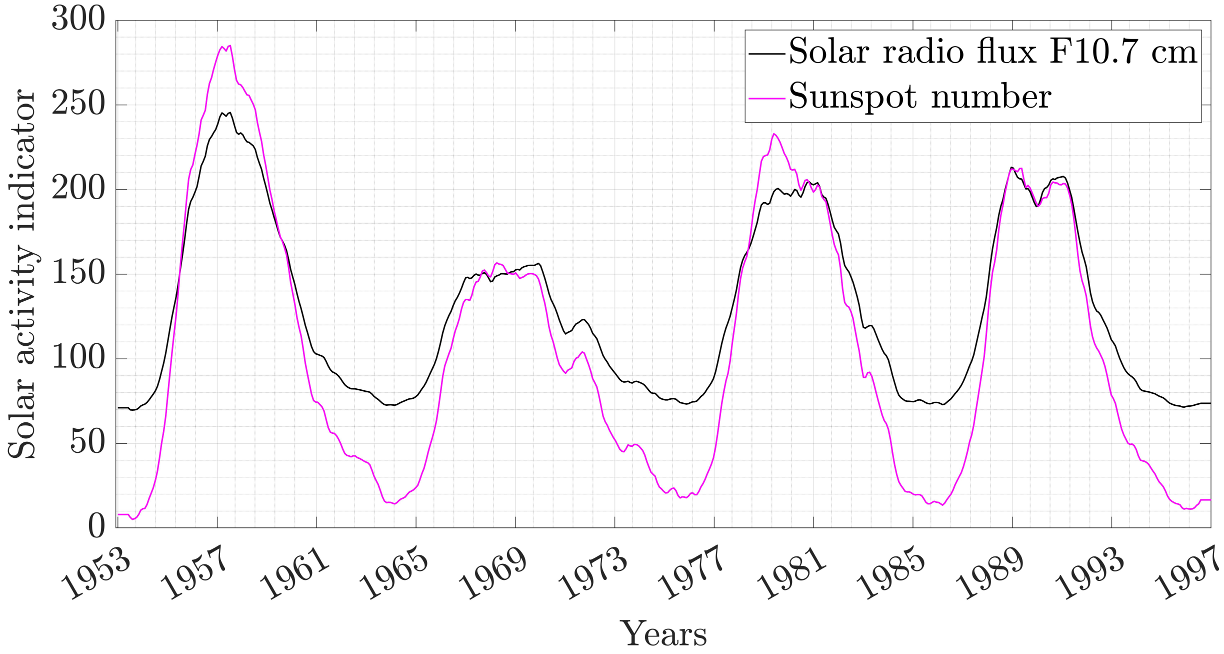

Figure 3: 13-Month Smoothed sunspot number and Solar radio flux FX10.7

Comparing **Figure 1** and **Figure 3**, it can be observed that the smoothed data became more useful for analysis, since some oscillatory components have been removed. In addition, it can be seen that the smoothed data does not have any delayed or too early predictions, making it a good forecast.

5 Multi-dimensional linear regression

A multi-dimensional regression has a following formula:

$$F_i = \beta_0 + \beta_1 R_i + \beta_2 R_i^2 + \beta_3 R_i^3 + \epsilon_i \tag{2}$$

The vector of monthly solar radio flux is the vector F of dependent variables. The matrix of independent variables R, regressors, contains values of sunspot numbers r_i at different times. Lastly, β is a vector of coefficients that can be derived from the *Least Squares Method.*

$$R = \begin{bmatrix} 1 & r_1 & r_1^2 & r_1^3 \\ 1 & r_2 & r_2^2 & r_2^3 \\ \dots & \dots & \dots & \dots \\ 1 & r_N & r_N^2 & r_N^3 \end{bmatrix} \quad F = \begin{bmatrix} f_1 \\ f_2 \\ \dots \\ f_N \end{bmatrix} \quad \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \dots \\ \beta_N \end{bmatrix} \tag{3}$$

Thus, the vector of coefficients by *LSM* can be determined as:

$$\beta = \left(R^T R\right)^{-1} R^T F \tag{4}$$

After calculations, for smoothed data the vector of coefficients is:

$$\beta = \left(66.4245;\ 0.0022 \cdot 10^5;\ -5.0329 \cdot 10^6\right) \tag{5}$$

6 Reconstruction of solar radio flux on the basis of sunspot number

For reconstruction of solar radio flux on the basis of sunspot number, **Equation 2** was used. The reconstruction was performed for unprocessed data and data after smoothing. The results obtained are shown in **Figure 4**.

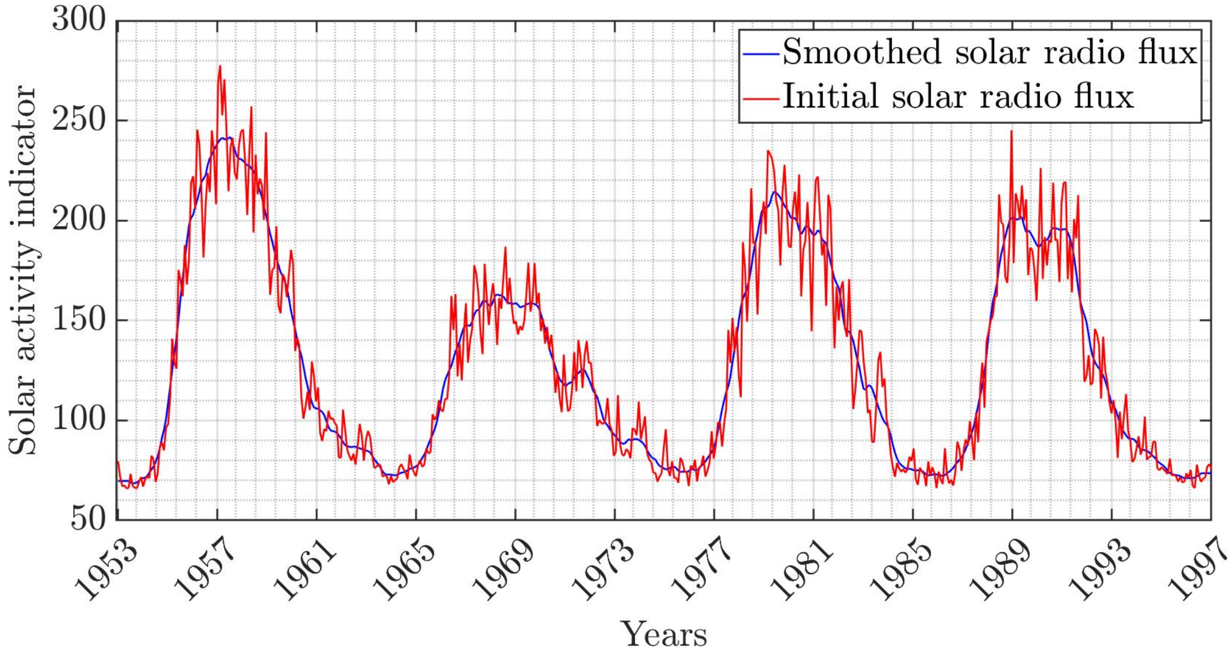

Figure 4: Smoothed and initial solar radio flux comparison

As can be seen, after the reconstruction using linear regression on unprocessed data, the final graph has a large oscillatory component. The variance of estimation error of solar radio flux was calculated with the following formula:

$$\sigma^2 = \frac{1}{N-1} \sum_{i=1}^{N} (f_i - \hat{f}_i)^2 \tag{6}$$

The results of calculations are displayed in the **Table 1**.

Variance	Result
$\sigma^2_{smoothed}$	24.4274
$\sigma^2_{initial}$	91.1115

Table 1: Variance of estimation for initial and smoothed cases

7 Conclusion

In this assignment the relationship between main indicators of solar activity such as sunspot number and the solar radio flux at 10.7 cm were studied.

Firstly, plots of initial data were built to provide a visual representation. The graphs showed that the initial data contains a large noise component which has to be addressed before analysis.

As well as that, scatter plot between monthly mean sunspot number and solar radio flux was provided. It showed the linear relationship between these two indicators. Sunspot numbers are the independent variables, with their increase the solar radio flux grows respectively.

We learnt new methods of dealing with noisy data. Particularly, we got familiar with 13th running mean that was applied to the sunspot number and solar radio flux. For the first and last six elements the arithmetic average was calculated, whereas for the middle values, the **Equation 1** was used.

By the end of the assignment, we managed to reconstruct the dynamics of the system from noisy data with the help of multi-dimensional linear regression. As the result shows, we achieved a relatively good forecasting as the prediction does not have any shifts, such as delays or early predictions. The value of variance for smoothed data is almost 4 times lower than for the raw data that shows mathematically that data contains less oscillations after processing.

We learnt how to extract a useful information from noise data, provide clear visual representation of the data in *Matlab*, implement 13-step running mean, applying the multi-dimensional linear regression and reconstructing the original data.

Individual contribution:

- Luca Breggion: Matlab code, report
- Irina Yareshko: Matlab code, report
- Ivan Kozhin: Matlab code, report