Skoltech

Skolkovo Institute of Science and Technology

Relationship Between Solar Radio Flux F10.7 and Sunspot Number

Experimental Data Processing

Assignment №1

Team №10

Submitted by:
Yunseok Park
Ilya Novikov
Ruslan Kalimullin

Instructor:
Tatiana Podladchikova

MA060238 (Term 1B, 2022-2023)

October 4th, 2022

Contents

1 Introduction

The objective of this Assignment is to apply multi-dimensional linear regression technique to find a connection between the primary indices of solar activity, sunspot number and solar radio flux at 10.7 cm (2800 MHz). This will lead to a better understanding of the least squares method (LSM), familiarity with the main indicators of solar activity and more detailed qualitative and quantitative understanding of important relationships.

2 Methodology

Question 3
The average monthly number of sunspots and the flux of solar radio emission F10.7cm is built for visual presentation. The plot is illustrated in Fig.1.

Question 4
Make a scatter plot between monthly mean sunspot number and solar radio flux F10.7 cm.

Question 5
Smoothing the monthly mean data (sunspot number and solar radio flux F10.7) by 13-month running mean. Formula for 13-month running mean is given in Equation 1.

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

Question 6 - 7
Construction of multi-dimensional regression by using Equation 2. Vector $F = |f_1 f_2 \cdots f_N|$ represents dependent variables(regressand). In this report F is the solar radio flux at 10.7 cm and $f_1 f_2 \cdots f_N$ represent the flux at different times.

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

$R = |1 r_1 r_1^2 r_1^3 1 r_2 r_2^2 r_2^3 \cdots 1 r_N r_N^2 r_N^3|$, $r_1, r_2, \cdots, r_N$ and $\beta = |\beta_0 \beta_1 \beta_2 \beta_3|$ represents matrix of independent variables(regressors), sunspot number at different times and vector of coefficients, respectively.

$$F = \begin{vmatrix} f_1 \\ f_2 \\ \cdots \\ f_N \end{vmatrix}, \qquad R = \begin{vmatrix} 1 & r_1 & r_1^2 & r_1^3 \\ 1 & r_2 & r_2^2 & r_2^3 \\ \cdots & \cdots & \cdots & \cdots \\ 1 & r_N & r_N^2 & r_N^3 \end{vmatrix}, \qquad \beta = \begin{vmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{vmatrix} \tag{3}$$

Question 8
Using LSM determine the vector of coefficients. The formula is given in Equation 4.

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

Question 9
Reconstruction of solar radio flux at 10.7 cm on basis of Equation 2.

Question 10
Variance of estimation error of solar radio flux F10.7 cm is determined by Equation 5.

$$\sigma^2 = \frac{1}{N-1} \sum_{i=1}^{N} (f_i - \bar{f}_i)^2 \tag{5}$$

3 Results

Question 3
Monthly mean sunspot number and solar radio flux F10.7 cm are plotted in the Figure 1. Since we used the raw data, sudden increases and decreases have been observed.

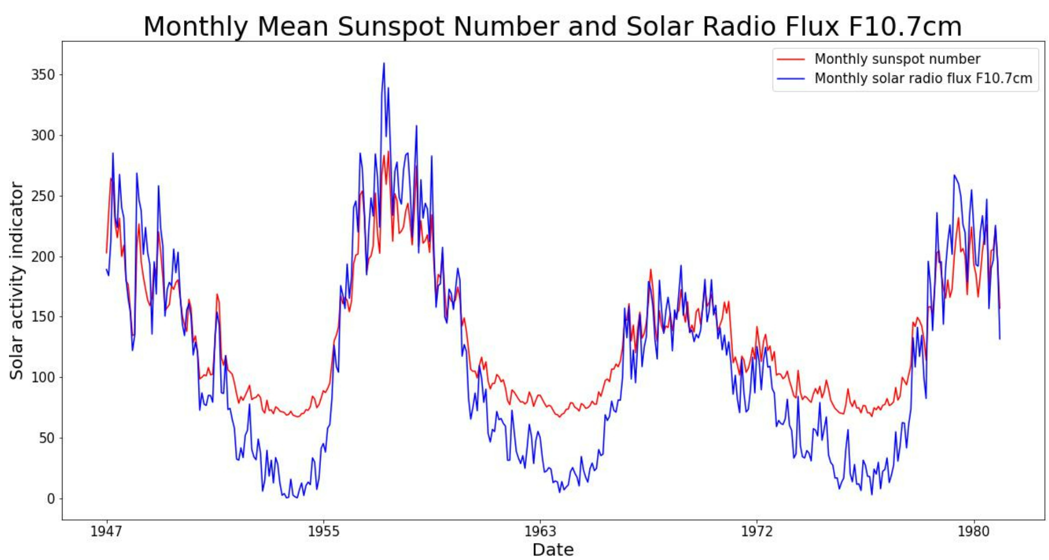


Figure 1: Monthly mean sunspot number and solar radio flux F10.7cm

Question 4
Scatter plot between monthly mean sunspot number and solar radio flux F10.7cm is plotted. A linear correlation between these two data can be observed. As the number of sunspots increases, the solar radio flux increases linearly.

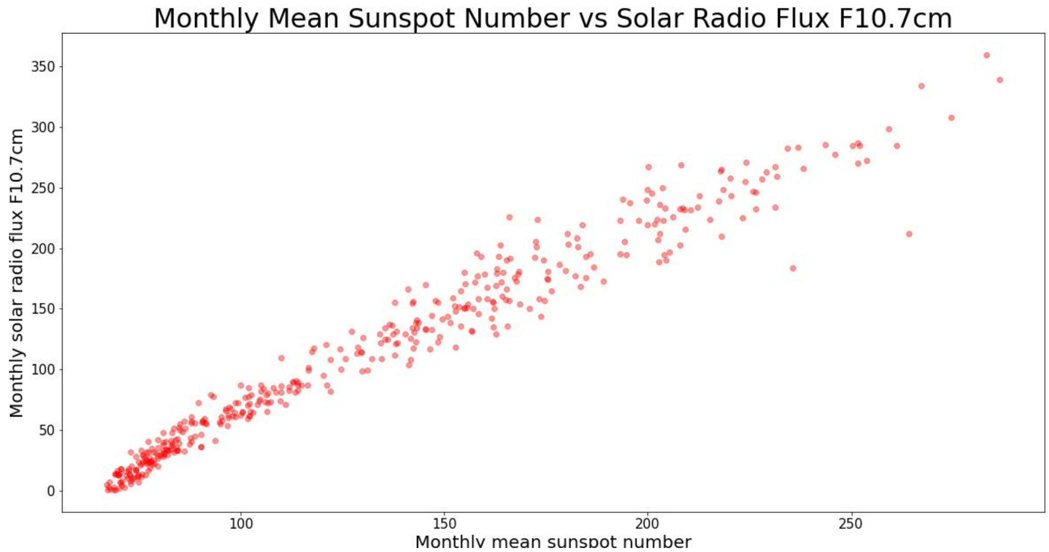


Figure 2: Monthly mean sunspot number vs solar radio flux F10.7cm

Question 5
To get rid of the noise from information smoothing of monthly mean sunspot number and solar radio flux F10.7cm is used. Abrupt jumps and sudden changes in the raw data have been eliminated via smoothing. As demonstrated it's possible predict a long-term forecast using monthly solar activity indicators.

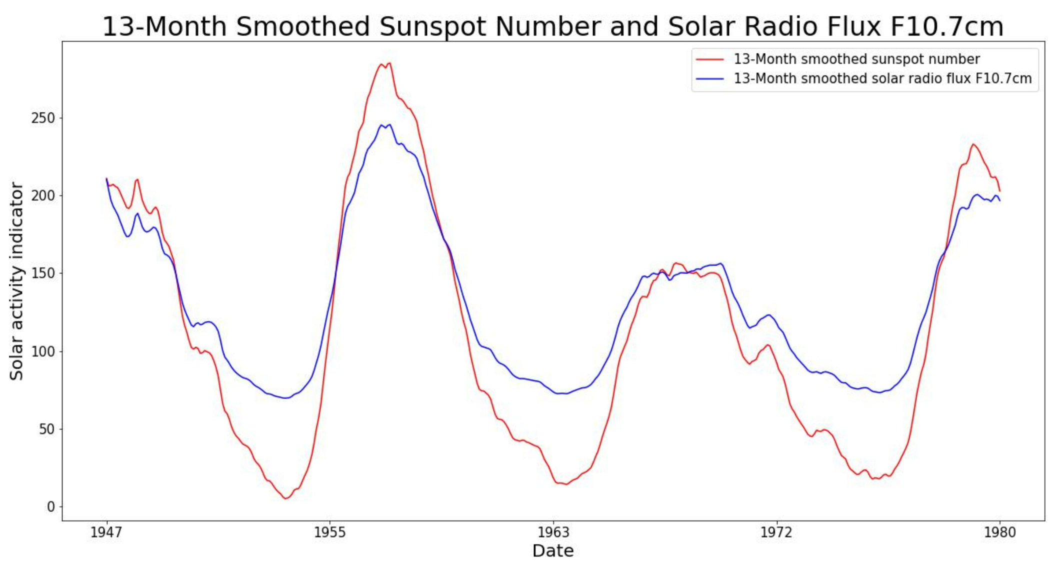


Figure 3: 13-month smoothed mean sunspot number and solar radio flux F10.7cm

In the Figure 3 raw and 13-month smoothed data are compared and we can clearly see that the noise is extracted from the original plot. The highest solar activity is observed in the early 60s. After reaching the highest solar activity at any time, a plateau occurs.

Question 6 - 8
After defining the regressands and the matrix of vectors the vector of coefficients is calculated:

β
$6.39221474e^{+1}$
$5.13574208e^{-1}$
$5.91876797e^{-4}$
$-5.55049581e^{-7}$

Question 9 - 10
After the reconstruction, variance has found as 12.77277717

$$\sigma^2 = \frac{1}{N-1}\sum_{i=1}^{N}(f_i - \bar{f}_i)^2 = 12.77277717 \tag{6}$$

After reconstruction we can also compared the smoothed solar radio flux with reconstructed flux. Comparison is demonstrated in Figure 4

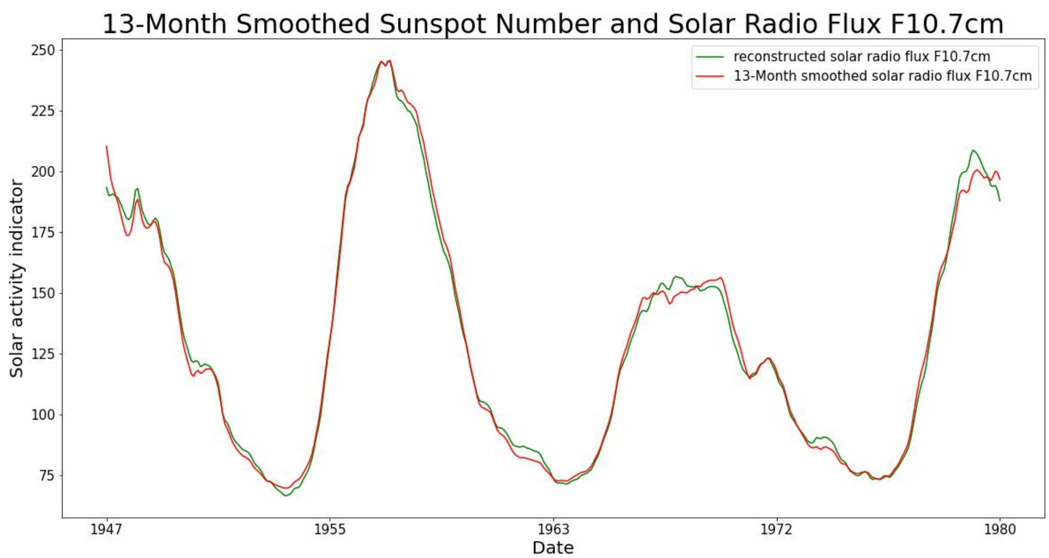


Figure 4: Smoothed solar radio flux vs reconstructed solar radio flux

4 Conclusion

On a final note, we plotted the sunspot number and the solar radio flux F10.7 to determine how noisy the raw data is. Next, we looked at the relationship between the number of sunspots and the flux of solar radio emission. WE have observed that as sunspot numbers (independent variable) increase solar radio flux (dependent variable) increase. Thus, looking at Figure 2, we can say that the solar flux increases as a dependent variable as the number of sunspots increases. Once we found the correlation of sunspot numbers and solar radio flux, we moved on to removing the noise from our raw data. We have learned how to extract useful information from raw data to forecast the relationship between sunspots and solar fluxes. To extract useful data, we used a 13-month running mean technique to smooth out the measurements. Smoothing is defined as a weighted average of noisy data. By this way, fluctuations in measurements are compensated by averaging. These fluctuations represent a change in dynamics that can lead to a significant increase in forecast errors. Therefore, we have isolated and smoothed out the patterns, as they allow us to study long-term forecasting. To smooth the data, we took the average of the first and last six values and determined that the 13-month raw data mean approaches the last value of the original matrix value. We have always used raw data for smoothing, not averages. In short, even though the dynamics are unknown, we reconstructed the process dynamics from the raw data to extract useful knowledge from the noise. In addition, in the running mean method, we don't have an optimal criterion and the change of mean is unknown. However, in the least squares method (LSM), we need to check the variance of the estimation error. It is also known that there is no other method that would be better for the linear case.

What we have learnt:

1. How to use numpy library for scientific computing in python
2. Develop multi-dimensional regression algorithms
3. Relation between solar radio flux and sunspot number
4. How to eliminate the noise via 13-month running mean
5. How the raw data gives larger variance whereas smoothed data results in smaller variance.

Contribution of members:

1. Ilya: wrote the code and plotted monthly mean sunspot number and solar radio flux.
2. Ruslan: wrote the code and plotted 13-month smoothing.
3. Yunseok: wrote the code for reconstruction and plotted the comparison of 13-month and reconstructed solar radio flux. Wrote the report.

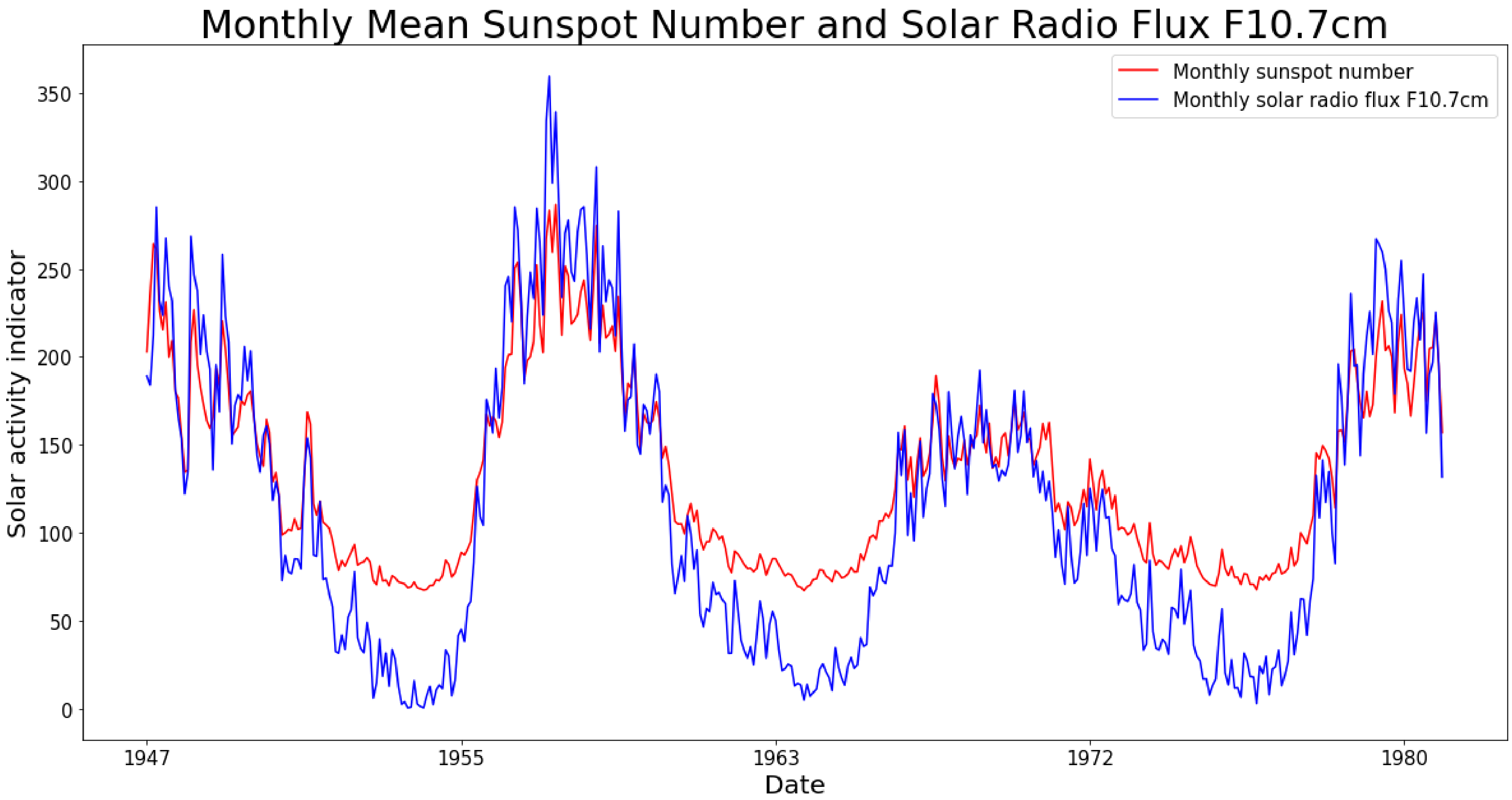
```
In [1]: import numpy as np
        import matplotlib.pyplot as plt
        import math
```

```
In [2]: data = np.loadtxt('data_group2.txt')
```

```
In [3]: years = data[:, 0]
        months = data[:, 1]
        fluxs = data[:, 2]
        sunspots = data[:, 3]
        count = len(months)
```

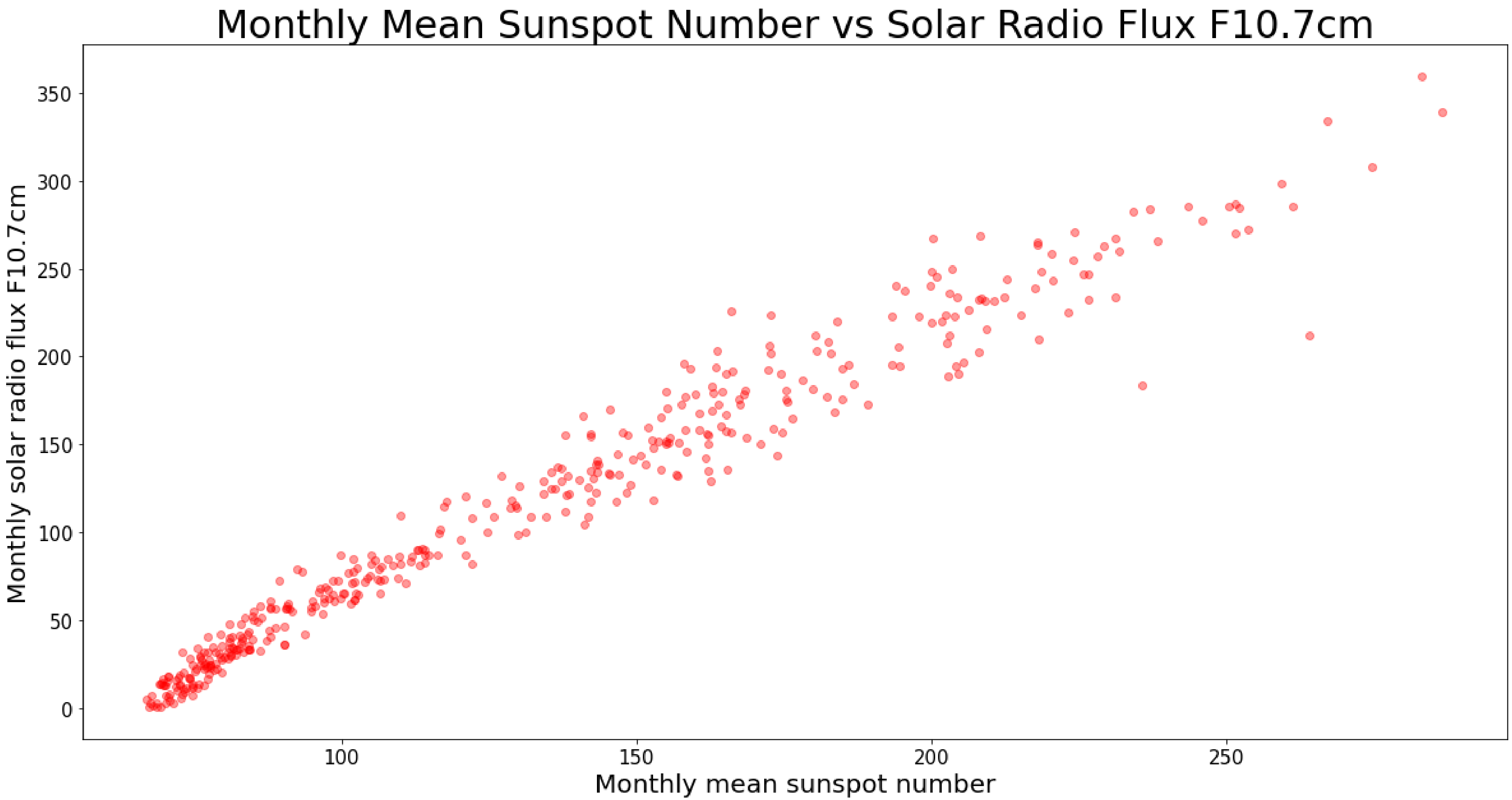
```
In [4]: #Smoothing of monthly mean data (sunspot number and solar radio flux F10.7) by 13-month running mean
        r_flux = np.zeros((count - 12, 1))  # smoothed fluxes array
        r_sunspots = np.zeros((count - 12, 1))  # smoothed sunspots array

        for n in range(r_flux.shape[0]):
            r_flux[n, 0] = 1/24 * (data[n, 2] + data[n + 12, 2]) + 1/12 * sum(data[n + 1:n + 12, 2])
            r_sunspots[n, 0] = 1/24 * (data[n, 3] + data[n + 12, 3]) + 1/12 * sum(data[n + 1:n + 12, 3])
```

```
In [5]: #Plot for monthly mean sunspot number and solar radio flux F10.7 cm
        fig, ax = plt.subplots(figsize=(20,10))
        ax.set_title('Monthly Mean Sunspot Number and Solar Radio Flux F10.7cm', fontsize = 30)
        ax.set_ylabel('Solar activity indicator', fontsize = 20)
        ax.set_xlabel('Date', fontsize = 20)
        ax.plot(fluxs, c='red', label='Monthly sunspot number')
        ax.plot(sunspots, c='blue', label='Monthly solar radio flux F10.7cm')
        ax.set_xticks([0, 100, 200, 300, 400], labels = ['1947', '1955', '1963', '1972', '1980'])
        ax.tick_params(axis='both', labelsize=15)
        ax.legend(fontsize = 15)
        plt.savefig("Monthly_sunspot_flux.jpg")
```

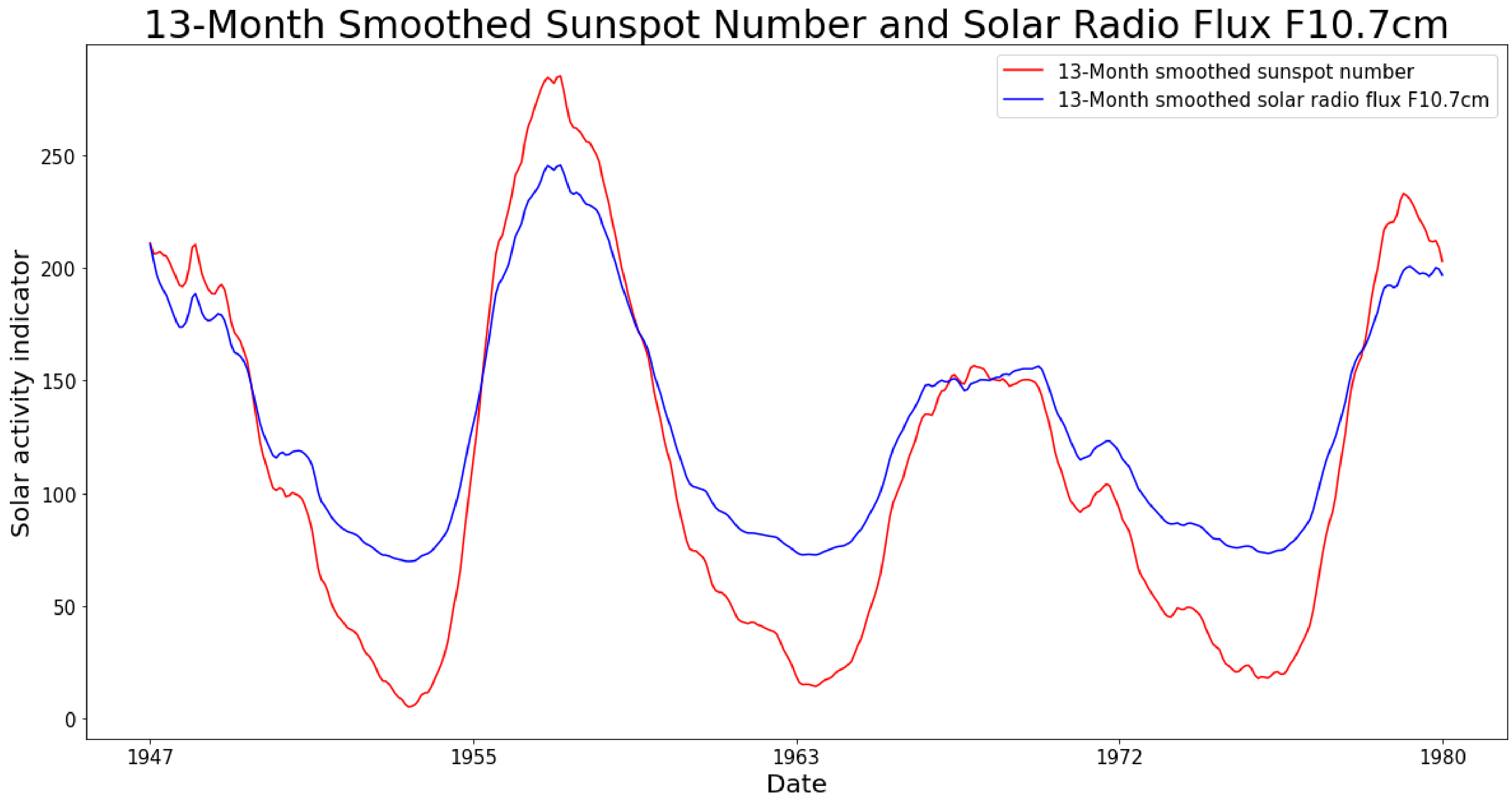


In [6]:

```
#Scatter plot between monthly mean sunspot number and solar radio flux F10.7 cm
fig, ay = plt.subplots(figsize=(20, 10))
ay.scatter(x=fluxs, y=sunspots, c='red', alpha=0.4)
plt.title('Monthly Mean Sunspot Number vs Solar Radio Flux F10.7cm', fontsize = 30)
plt.xlabel('Monthly mean sunspot number', fontsize = 20)
plt.ylabel('Monthly solar radio flux F10.7cm', fontsize = 20)
ay.tick_params(axis='both', labelsize=15)
plt.savefig("Sunspot_vs_flux.jpg")
```



In [7]:
```
#Plot for monthly mean data (sunspot number and solar radio flux F10.7) by 13-month running mean
fig, az = plt.subplots(figsize=(20,10))
az.set_title('13-Month Smoothed Sunspot Number and Solar Radio Flux F10.7cm', fontsize = 30)
az.set_ylabel('Solar activity indicator', fontsize = 20)
az.set_xlabel('Date', fontsize = 20)
az.plot(r_sunspots, c='red', label='13-Month smoothed sunspot number')
az.plot(r_flux, c='blue', label='13-Month smoothed solar radio flux F10.7cm')
az.set_xticks([0, 100, 200, 300, 400], labels = ['1947', '1955', '1963', '1972', '1980'])
az.tick_params(axis='both', labelsize=15)
az.legend(fontsize = 15)
plt.savefig("13_smoothed.jpg")
```



In [8]:
```
#Determination of vector of regressands and matrix of regressors vector
F = r_flux
R = np.ones((len(F), 4))
R[:, 1] = r_sunspots[:, 0]
R[:, 2] = r_sunspots[:, 0] **2
R[:, 3] = r_sunspots[:, 0] **3
```

In [9]:
```
print('Matrix of regressors vector: \n', R, '\n')

#Determination of vector of coefficients by LSM
B1 = np.transpose(R).dot(R)
B2 = np.linalg.inv(B1)
```

```python
B3 = B2.dot(np.transpose(R))
B = B3.dot(F)
```

```
Matrix of regressors vector:
 [[1.00000000e+00 2.10895833e+02 4.44770525e+04 9.38002505e+06]
 [1.00000000e+00 2.06045833e+02 4.24548854e+04 8.74765225e+06]
 [1.00000000e+00 2.06333333e+02 4.25734444e+04 8.78432070e+06]
 ...
 [1.00000000e+00 2.11929167e+02 4.49139717e+04 9.51858059e+06]
 [1.00000000e+00 2.09141667e+02 4.37402367e+04 9.14790601e+06]
 [1.00000000e+00 2.02845833e+02 4.11464321e+04 8.34638231e+06]]
```

In [10]:

```python
#Reconstruct solar radio flux F10.7 on the basis of sunspot number
F1 = R.dot(B)
```

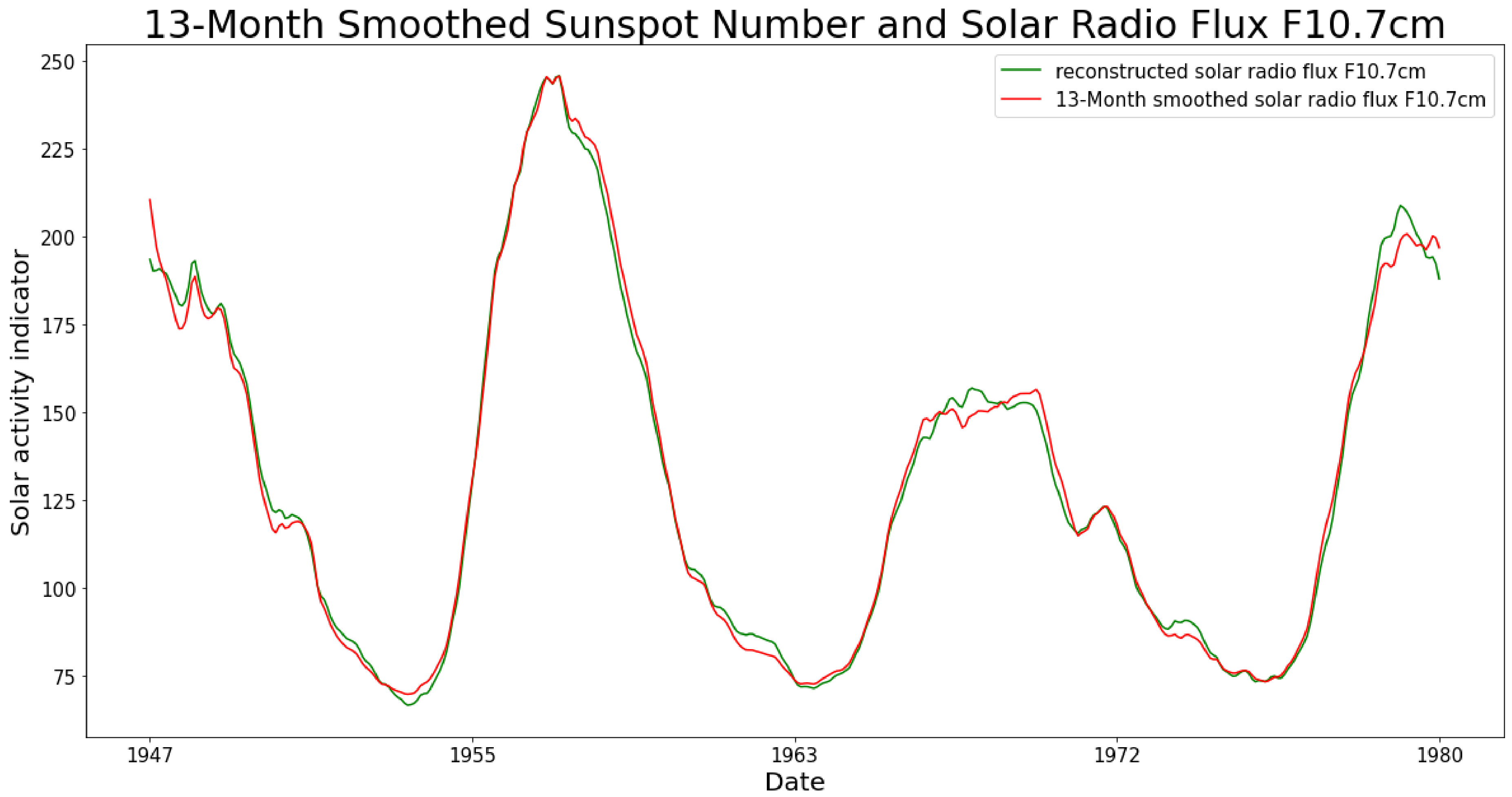
In [11]:

```python
#Determine the variance of estimation error of solar radio flux F10.7
delta = sum(np.power((F - F1), 2)) / (400)
print('Varience of estimation error: \n', round(delta[0], 4))
```

```
Varience of estimation error:
 12.7728
```

In [12]:

```python
#Comparison with reconstructed and smoothed
fig, ar = plt.subplots(figsize=(20,10))
ar.set_title('13-Month Smoothed Sunspot Number and Solar Radio Flux F10.7cm', fontsize = 30)
ar.set_ylabel('Solar activity indicator', fontsize = 20)
ar.set_xlabel('Date', fontsize = 20)
ar.plot(F1, c='green', label='reconstructed solar radio flux F10.7cm')
ar.plot(r_flux, c='red', label='13-Month smoothed solar radio flux F10.7cm')
ar.set_xticks([0, 100, 200, 300, 400], labels = ['1947', '1955', '1963', '1972', '1980'])
ar.tick_params(axis='both', labelsize=15)
ar.legend(fontsize = 15)
plt.savefig("reconstructed.jpg")
```



13-Month Smoothed Sunspot Number and Solar Radio Flux F10.7cm
reconstructed solar radio flux F10.7cm
13-Month smoothed solar radio flux F10.7cm
Solar activity indicator
250
225
200
175
150
125
100
75
1947
1955
1963
1972
1980
Date