Assignment 1

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

by:

Ahmed Baza, Skolteck, 2020

Ahmed Elgohary, Skolteck, 2020

Mohamed Sayed, Skolteck, 2020

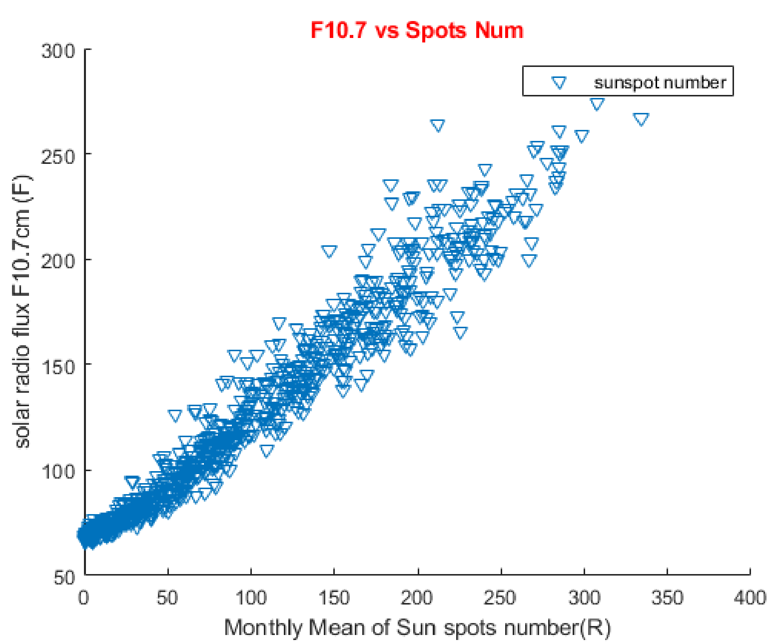
Ziad Baraka, Skolteck, 2020

1-Load monthly mean sunspot number and solar radio flux F10.7cm measurements to MATLAB:

```
clc
clear
load('data_group1.mat');
Flux=data_group1(:,3); %monthly solar radio flux F10.7cm
Spots=data_group1(:,4); %monthly mean sunspot number
Years_axis=data_group1(:,1)+(1/12)*data_group1(:,2); %An array for ploting data (years seperated by 12 points or months)
```

2-Makeing a scatter plot between monthly mean sunspot number and solar radio flux F10.7cm.

```
scatter(Spots,Flux,'v');
xlabel('Monthly Mean of Sun spots number(R)');
ylabel('solar radio flux F10.7cm (F)');
title('F10.7 vs Spots Num','Color','r');
legend('sunspot number');
```



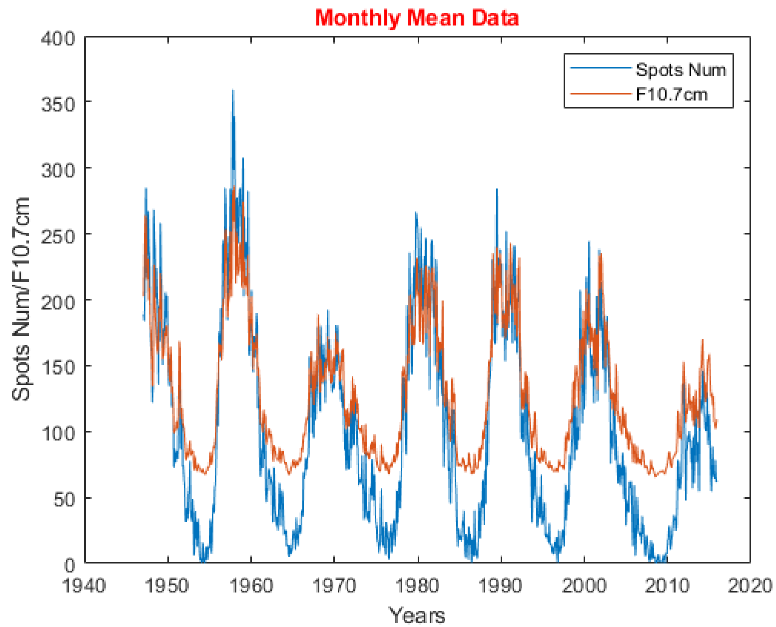
[figure(1): monthly average of F10.7 flux density Vs. Spots Number]

Conclusion:

Looking at the scatter plot we can clearly see a direct correlation between the radio flux and the mean of the number of sunspots in a month, which means we can rely on the sunspot number as an indicator for estimating the radio flux 10.7cm prediction and analysis purposes.

```
% ploting the Data befor smoothing:
```

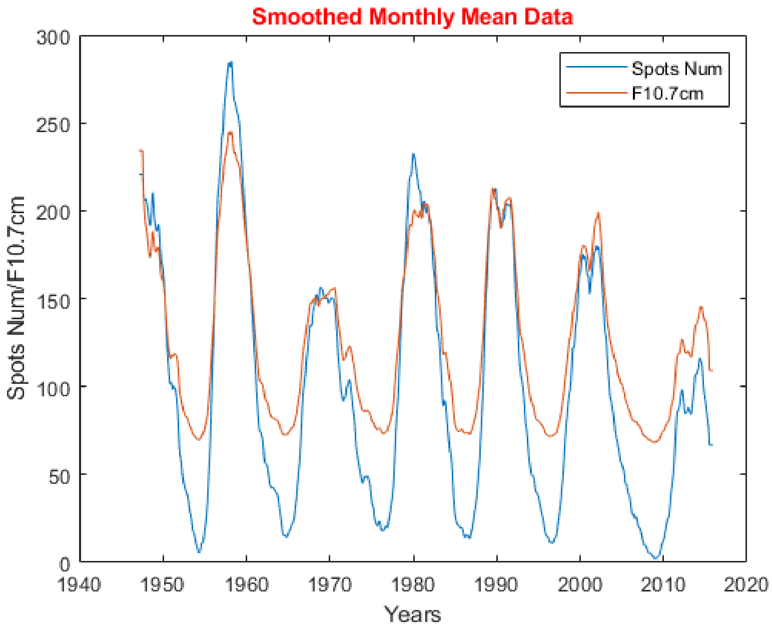
```
plot(Years_axis,Spots,Years_axis,Flux)
title('Monthly Mean Data','Color','r');
xlabel('Years');
ylabel('Spots Num/F10.7cm');
legend('Spots Num','F10.7cm');
```



[Figuer(2): Data befor smooting]

3-Make smoothing of monthly mean data (sunspot number and solar radio flux F10.7) by 13-month running mean. Plot results.

```
N=length(Flux); % number of data points
FS=1:N; SS=1:N; %array for the smoothing results.(FS:: Flux smoothed, SS:: Spots smoothed)
for i=7:N-6 %somthing
    FS(i)=(1/24)*(Flux(i-6)+Flux(i+6))+(1/12)*sum(Flux(i-5:i+5));
    SS(i)=(1/24)*(Spots(i-6)+Spots(i+6))+(1/12)*sum(Spots(i-5:i+5));
end
FS(1:6)=mean(Flux(1:6)); FS(N-5:N)=mean(Flux(N-5:N));
SS(1:6)=mean(Spots(1:6)); SS(N-5:N)=mean(Spots(N-5:N));
plot(Years_axis,SS,Years_axis,FS);
title('Smoothed Monthly Mean Data','Color','r');
xlabel('Years');
ylabel('Spots Num/F10.7cm');
legend('Spots Num','F10.7cm');
```

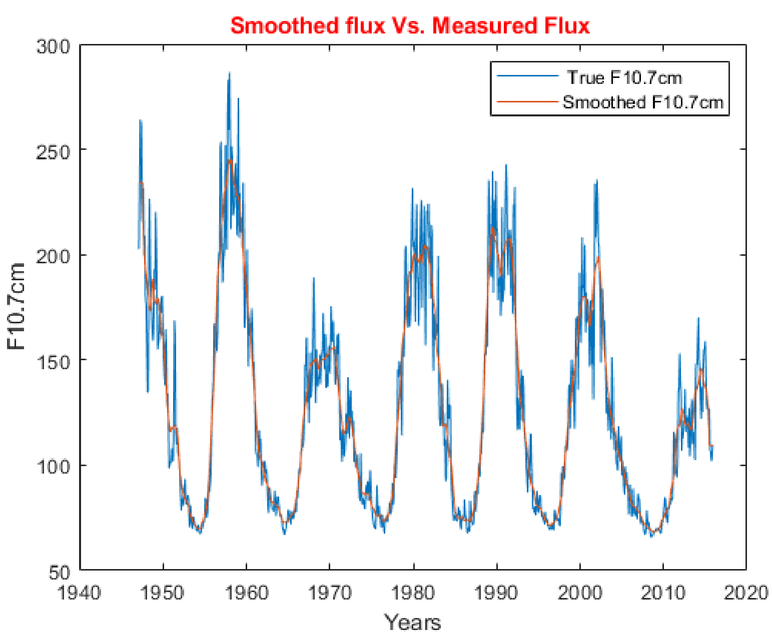


[Figuer(3): Data After Smoothing]

Comment:

Plotting both the raw data and the smoothed data clearly shows a big difference between the two lines, the raw data graph is too noisy to reach a conclusion or make any dependable correlation, whilst the smoothed graph makes it clearer for a data analyst to build up models and reach more reliable conclusions. Also, The 13-months smoothing seems like a good fit as it didn't lose the annual trend for data, but it might have lost some of the monthly features of the raw data. Hence, if we needed to analyze data on a narrower range of time a 13-months smoothing would be inefficient but as for the annual prediction it looks like a good approximation.

```
plot(Years_axis,Flux,Years_axis,FS);
title('Smoothed flux Vs. Measured Flux ','Color','r');
xlabel('Years');
ylabel('F10.7cm');
legend(' True F10.7cm','Smoothed F10.7cm');
```



[Figuer(4): F10.7Cm Befor & After Smoothing]

comment:

To see the difference between the raw and the smoothed data both were plotted simultaneously. And we can clearly read from the figure that the smoothed data is following the general trend of the raw data but without the noise, and that the values are also close.

4-Construction of multi-dimensional linear regression.

```
RM=[ones(N,1),SS', SS.^2', SS.^3']; % constracting the R matrix (independant variables) regressors.
% the monthly solar radio flux F10.7cm is the regressand (dependent variable).
% let's call the vector of coefficients [B].
```

5-Determine vector of coefficients by LSM.

```
B=(RM'*RM)\(RM'*FS');
disp('Vector of coefficients:');
```

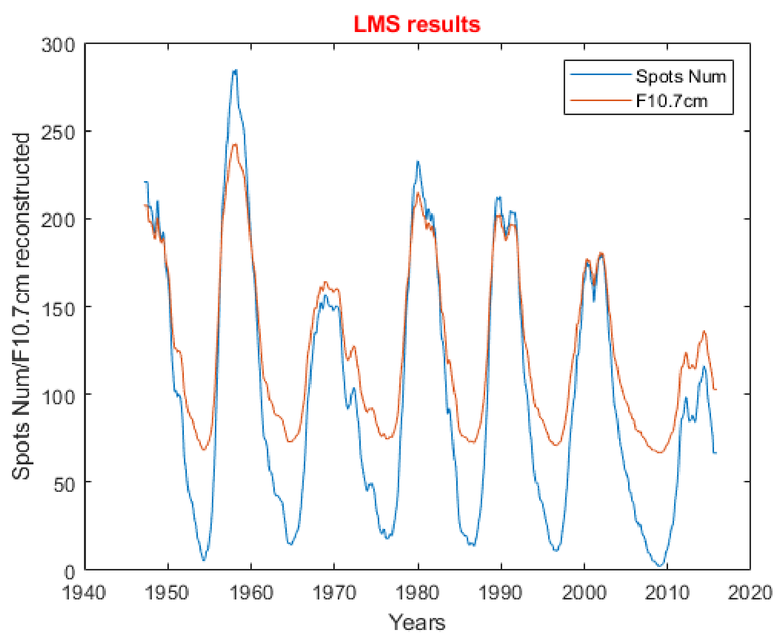
```
Vector of coefficients:
```

```
disp(B);
```

```
  65.7969
   0.4619
   0.0017
  -0.0000
```

6-Reconstruct solar radio flux at 10.7 cm on the basis of sunspot number and plot.

```
FR=RM*B; %reconstructed Solar Flux F10.7cm.
```

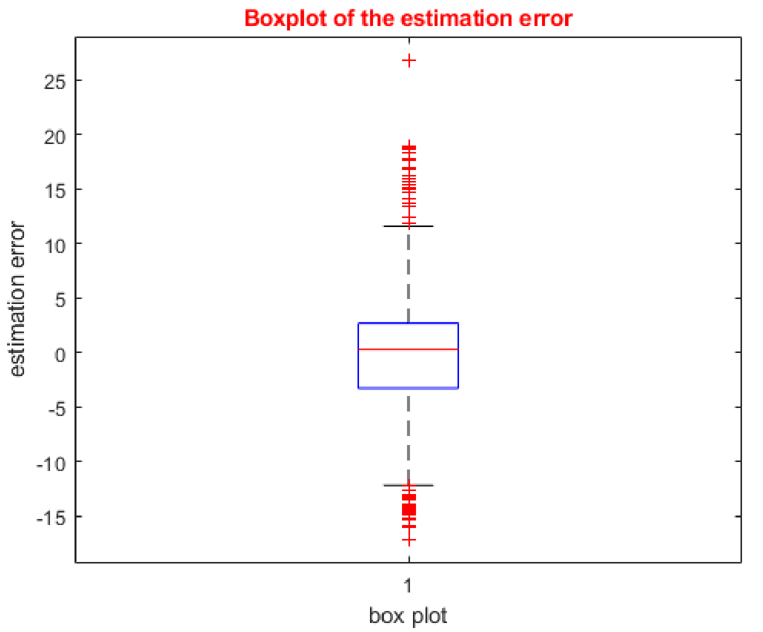


7-Determine the variance of estimation error of solar radio flux at 10.7.

```
Variance= (1/(N-1))*sum((FS'-FR).^2);
fprintf('%s =Variance:',Variance);
```

```
3.833184e+01 Variance .
```

```
Fdiff=FS-FR';
minD=max(Fdiff);
maxD=min(Fdiff);
standard=sqrt(Variance);
boxplot(Fdiff)
title('Boxplot of the estimation error','Color','r')
ylabel('estimation error')
xlabel('box plot')
```



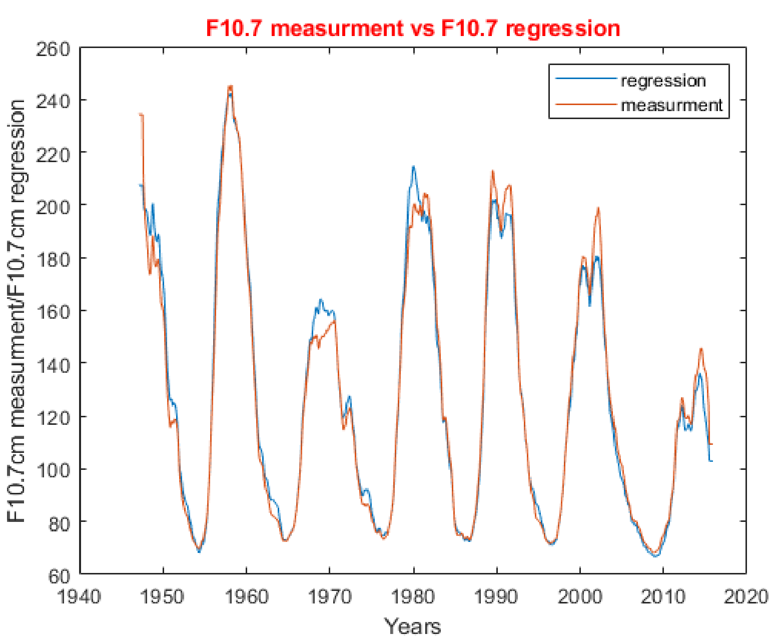
[Figuer(5): Boxplot of estimation error]

Comment:

To make sure the variance value is good we used the standard deviation and box plot. As we studied in probability before, a good value of standard deviation is which 98% of the values are enclosed within 3 times the standard deviation for a normal distribution of the data, and after calculation the standard deviation was equal to 6.19 and the maximum value and minimum value of the data were 18.7 and -17.15 excluding the first and last six points which were averaged for smoothing. And the boxplot showed almost no skew and the values were around the median is 0.3 estimation error value which is so close to the perfect estimation error of zero percent.

8- Measured data vs linear regression model.

```
plot(Years_axis,FR,Years_axis,FS)
title('F10.7 measurment vs F10.7 regression','Color','r');
xlabel('Years');
ylabel('F10.7cm measurment/F10.7cm regression');
legend('regression','measurment')
```



[Figuer(6): Measured Flux and LMS constructed]

Comment:

The reconstruction of the radio flux would appear at the beginning closely similar to the data, but we can identify clearly a difference between the values in most of the peaks. The model was predicting higher values whilst the data showed a lower value and vice versa.

Learning Log:

A) Reflecting on the scatter plot:

we can find almost a linear correlation between the monthly mean sunspots number and the solar flux 10.7 cm, based on that we decided to use a first order model and compare it with the 3rd order model assigned in the assignment.

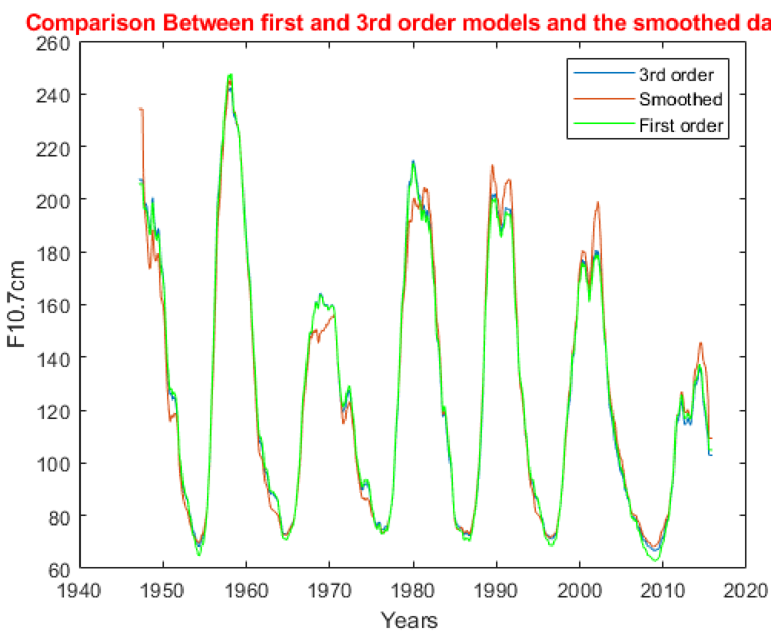
Comparing the coefficients values and variance we found:

1- β0, β1 in our model is close to β0, β1 in the 3rd order mode and β2, β3

2- The variance didn't increase greatly. Using the method in the assignment it was almost 38 whilst using the first order approximation it was almost 41.

So we concluded that a first order model can be applied to the data with acceptable accuracy.

```
RM1=[ones(N,1),SS'];
B1=(RM1'*RM1)\(RM1'*FS');
FR1=RM1*B1;
plot(Years_axis,FR,Years_axis,FS,Years_axis,FR1,'g')
title('Comparison Between first and 3rd order models and the smoothed data ','Color','r')
xlabel('Years ')
ylabel('F10.7cm')
legend({'3rd order','Smoothed','First order'})
```




```
Variance1= (1/(N-1))*sum((FS'-FR1).^2)
```

```
Variance1 = 41.1706
```

B) Reflecting on the LMS method:

We understood more clearly how to apply the LSM method and have a practical sense for numbers along with the values of models based on the LSM. However, it's not a magical tool that you plug in any data and expect good approximations. You have to understand the relations between the variables and the dynamics of the problem to be able to assess the validity of your estimation and what can be acceptable. And if we are giving a general bullet points for a conclusion that we can reflect upon in future calculations, we believe they would be:

- That data cannot mean anything unless we can understand it and formulate relations between variables, that's why we need smoothing for practical approximations.
- That smoothing depends on the range of time you want to study, if it's a big range of dependent variables, then you can use a big range of independent variables and vice versa.

For example: in this assignment if we were concerned about the monthly trends then using the 13 months smoothing would lose some of the features but a 7-months would be fitter to Use.

- Using the fittest method for our problem might also have some wrong estimations, but it would be the best approximation we can come up with. Like what happened in some peaks, that the model either estimated higher or lower values for the flux. (you should always expect some errors)
- If the least square method was the one that fits our physical problem the most, then it is reliable for some predictions, as the model at the end provided a good estimate for the flux