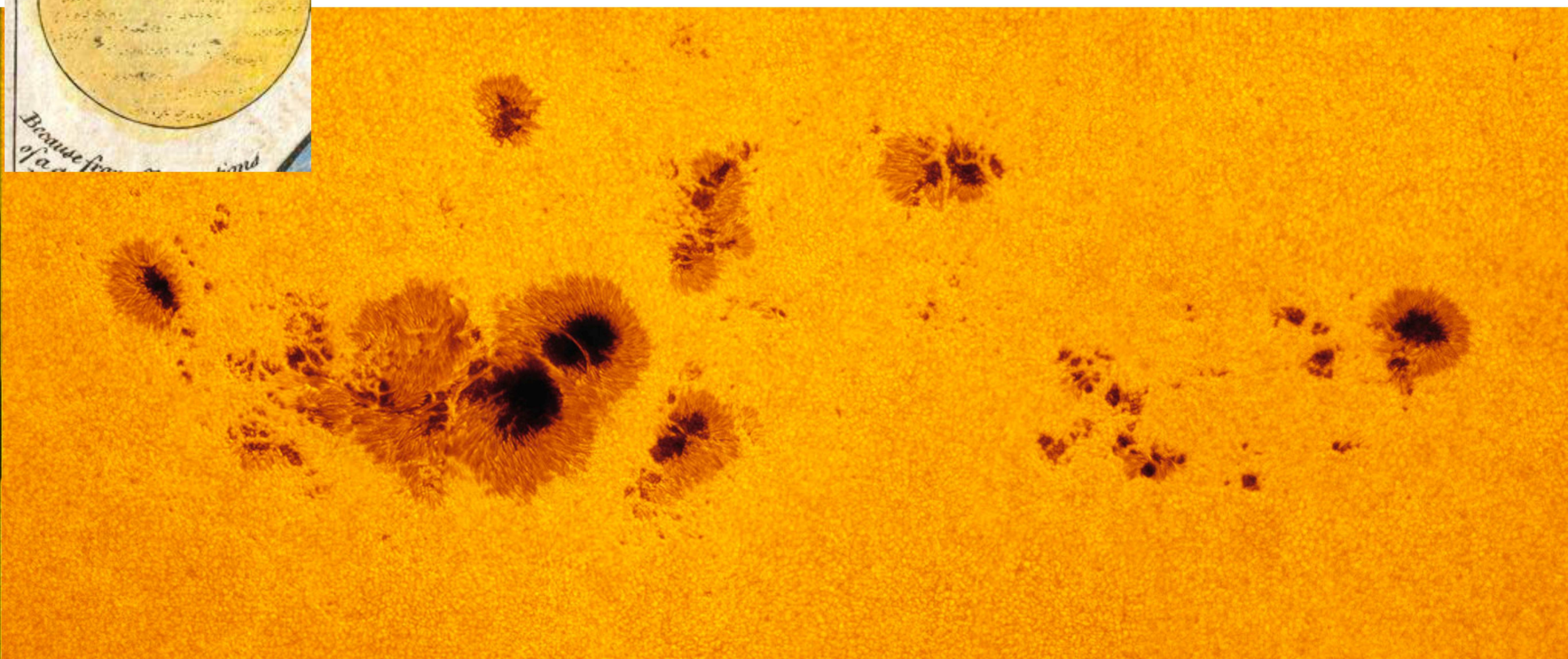
Homework #2: Plotting number of sunspots vs time and calculating the "running" or "moving" average

1794 observation

Modern observation



Homework #2: Plotting number of sunspots vs time and calculating the “running” or “moving” average

This is exercise 3.1 in the Newman textbook

Exercise 3.1: Plotting experimental data

In the on-line resources[3] you will find a file called **`sunspots.txt`**, which contains the observed number of sunspots on the Sun for each month since January 1749. The file contains two columns of numbers, the first being the month and the second being the sunspot number.

a) Write a program that reads in the data and makes a graph of sunspots as a function of time.

b) Modify your program to display only the first 1000 data points on the graph.

c) Modify your program further to calculate and plot the running average of the data, defined by

$$Y_k = \frac{1}{2r+1} \sum_{m=-r}^{r} y_{k+m}$$

where $r = 5$ in this case (and the y_k are the sunspot numbers). Have the program plot both the original data and the running average on the same graph, again over the range covered by the first 1000 data points.

```
0	58.0
1	62.6
2	70.0
3	55.7
4	85.0
5	83.5
6	94.8
7	66.3
8	75.9
9	75.5
10	158.6
11	85.2
12	73.3
13	75.9
14	89.2
15	88.3
16	90.0
17	100.0
18	85.4
19	103.0
20	91.2
21	65.7
```

Sunspot data found here:
http://www-personal.umich.edu/~mejn/cp/data/sunspots.txt

Each line contains two numbers separated by a tab.

The first number is the month (counting months since January 1749)

The 2nd number is the number of observed sunspots.

We can read in this data using numpy loadtxt().

This will return a 2 x 3143 numpy array. You can use slicing to split this into two arrays.

(see Notebook 05 in the scatter plot section for an example of reading in a different dataset [star magnitude vs temperature] or see Page 92 or Page 100 of the textbook)

Once we have two numpy arrays, each with 3143 entries, we can plot them.

a) Plot the month on the x-axis and the number of sunspots on the y-axis.

b) Make the plot again, but this time only plot the first 1000 months. Hint: Use slicing.

c) Finally (the hard part): calculate the moving average for this data. Plot the moving average on top of the month data (for all of the data and for the first 1000 months)

“Running” or “Moving” average

Common technique to smooth data

"Running" or "Moving" average

Index

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19

Data array

9	6	8	4	9	7	5	8	5	3	5	7	5	4	3	4	2	4	1	4

r = 5 (parameter given in problem)

for any given index i in the array, we want to calculate the average of the values within the window [i-r, i+r]. We will then save that value within a new array.

In order to fit the entire window we should start at index 5 and finish at the 5th to last index

“Running” or “Moving” average

Index

0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**

Data array

9	**6**	**8**	**4**	**9**	**7**	**5**	**8**	**5**	**3**	**5**	**7**	**5**	**4**	**3**	**4**	**2**	**4**	**1**	**4**

Start at index 5, calculate the average
within the window [0,10]
= (sum)/(window size)
= (9+6+8+4+9+7+5+8+5+3+5)/11
= 6.27

Moving average list/array

6.27									

“Running” or “Moving” average

Index

0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**

Data array

9	**6**	**8**	**4**	**9**	**7**	**5**	**8**	**5**	**3**	**5**	**7**	**5**	**4**	**3**	**4**	**2**	**4**	**1**	**4**

Iterate to index 6, calculate the average
within the window [1,11]
= (6+8+4+9+7+5+8+5+3+5+7)/11
= 6.09

Moving average list/array

6.27	**6.09**								

“Running” or “Moving” average

Index

0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**

Data array

9	**6**	**8**	**4**	**9**	**7**	**5**	**8**	**5**	**3**	**5**	**7**	**5**	**4**	**3**	**4**	**2**	**4**	**1**	**4**

Iterate to index 7, calculate the average
within the window [2,12] = 6.09

Moving average list/array

6.27	**6.09**	**6.0**							

“Running” or “Moving” average

Index

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19

Data array

9	6	8	4	9	7	5	8	5	3	5	7	5	4	3	4	2	4	1	4

The last window is not centered on the last element of the array.

Moving average list/array

6.27	6.09	6.0							3.81

“Running” or “Moving” average

Index

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19

Data array

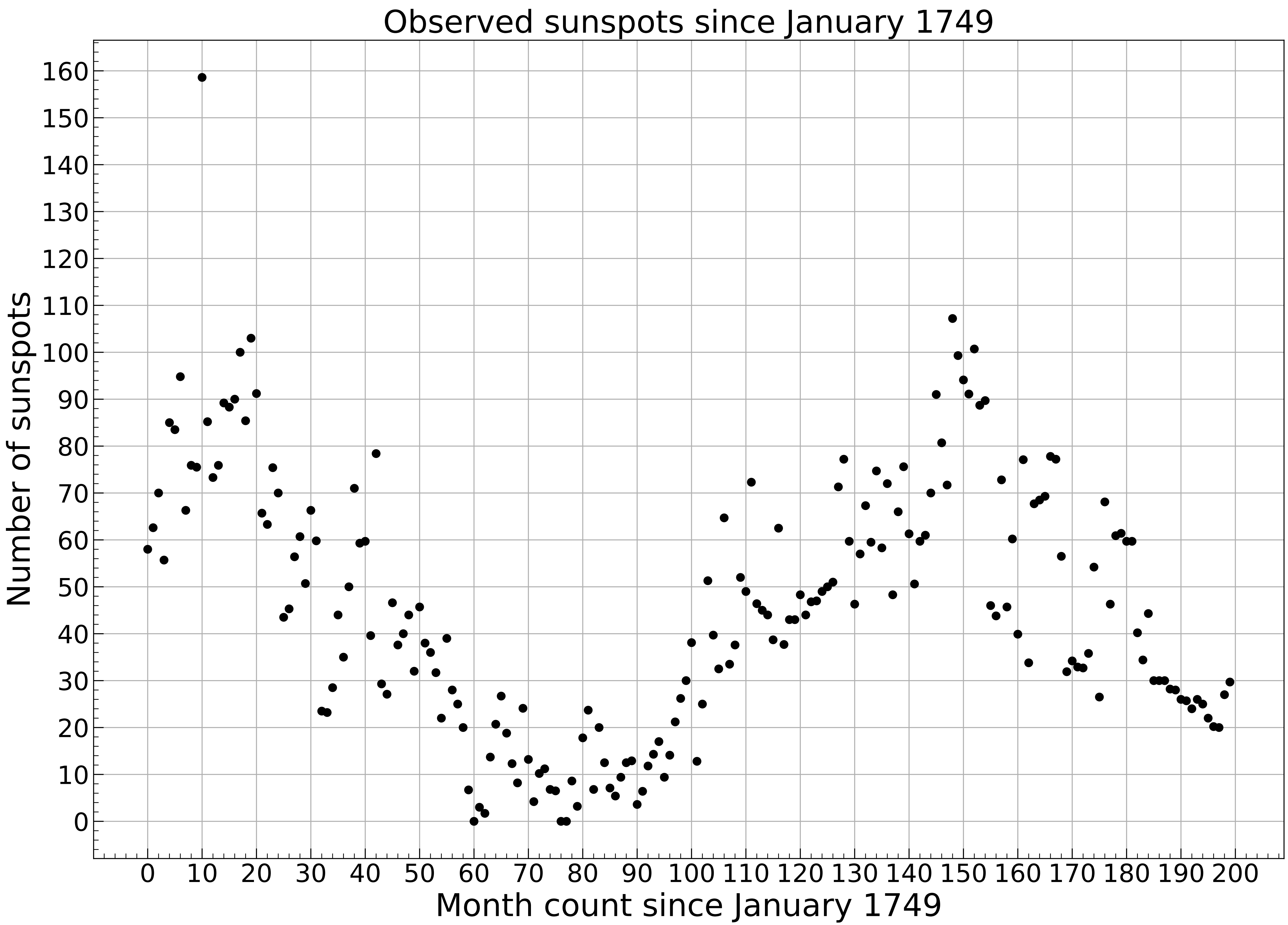
9	6	8	4	9	7	5	8	5	3	5	7	5	4	3	4	2	4	1	4

Note: In this example we are considering 20 months of data but our moving average list only contains 10 entries (r=5, so the first window is centered on index 5 of the month and data array). To plot this correctly you will have to use slicing to decrease the size of the month array.

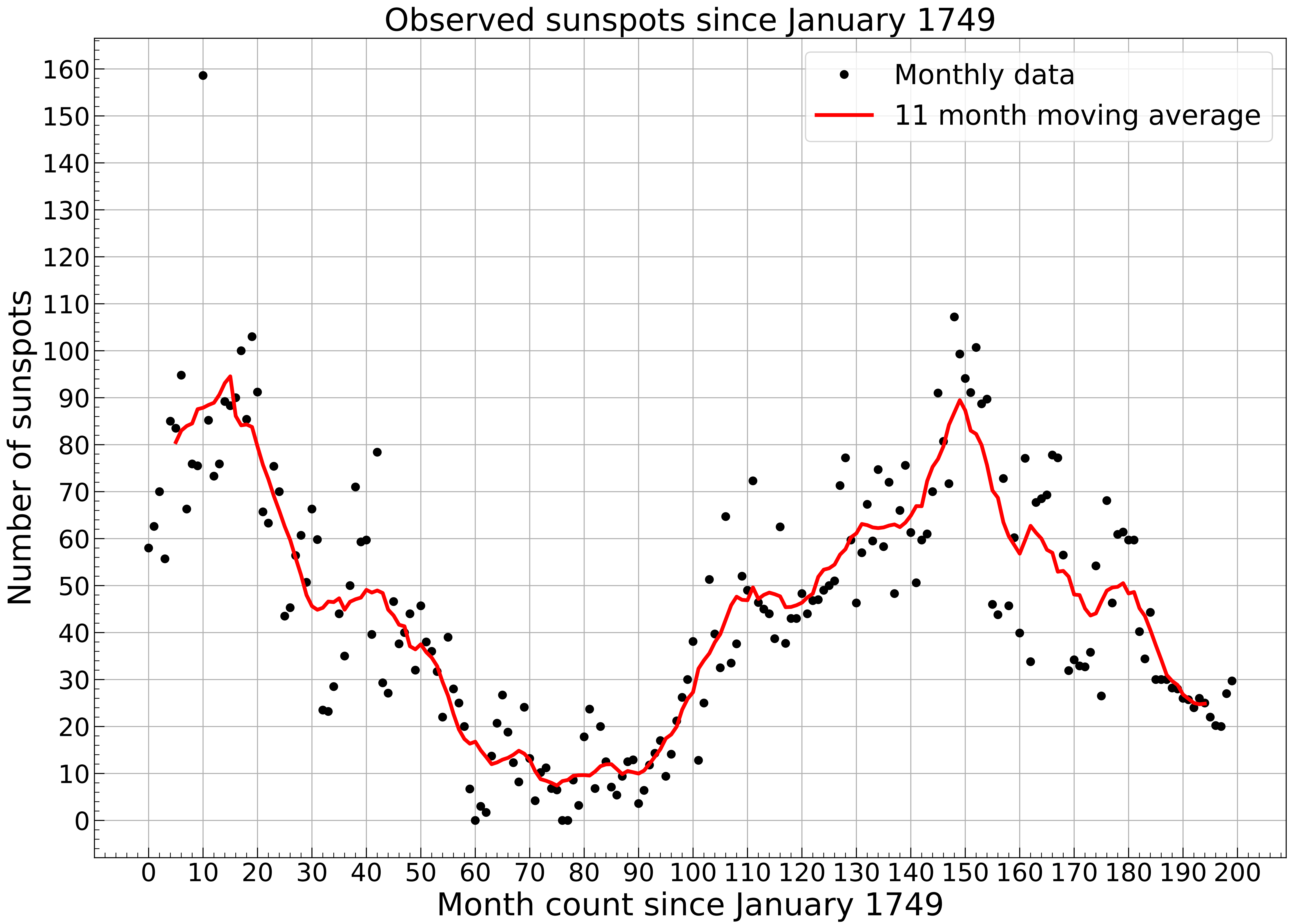
Moving average list/array

6.27	6.09	6.0							3.81

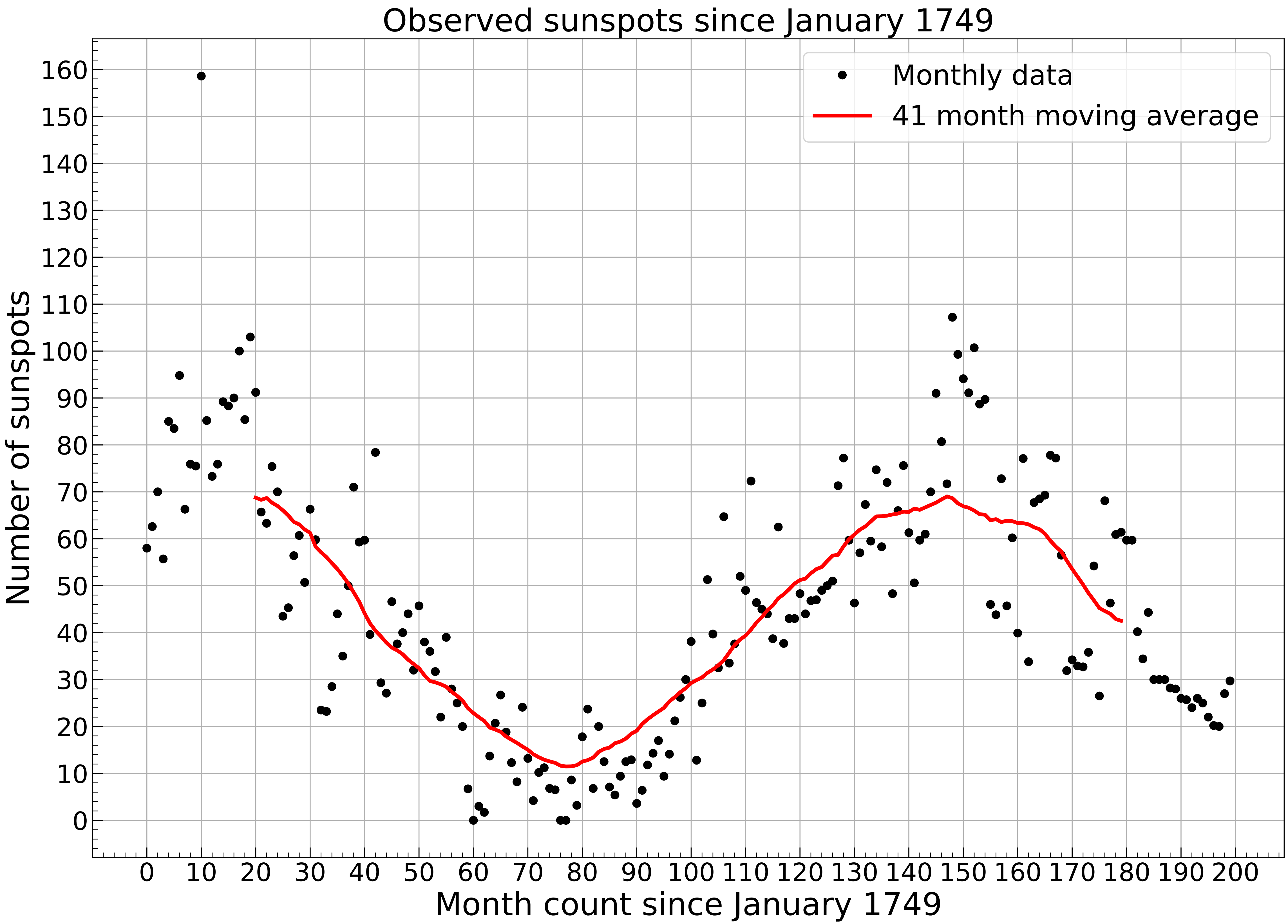
Example plot (200 months shown here. You will plot 1000 months and all months)



Example plot (200 months shown here. You will plot 1000 months and all months). Here r = 5.



Example plot (200 months shown here. You will plot 1000 months and all months). Here r = 20.



Example plot (200 months shown here. You will plot 1000 months and all months). Here r = 50.