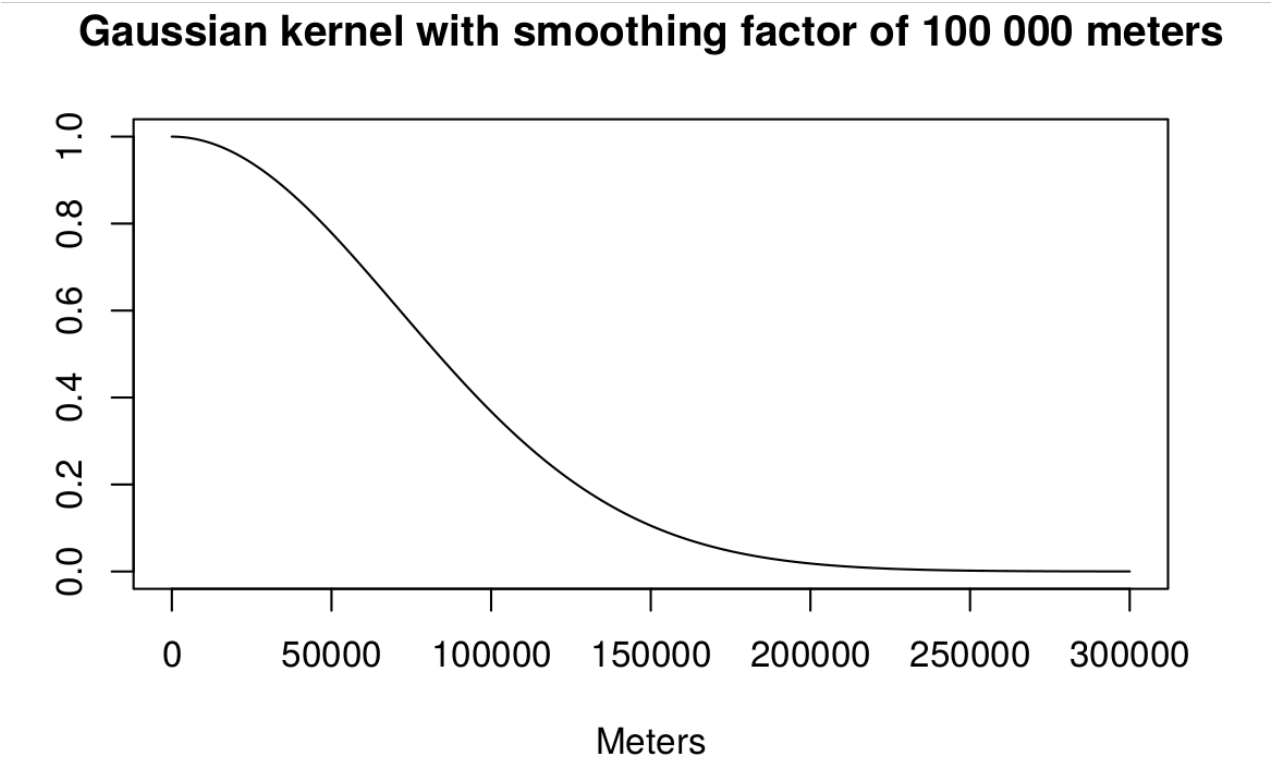
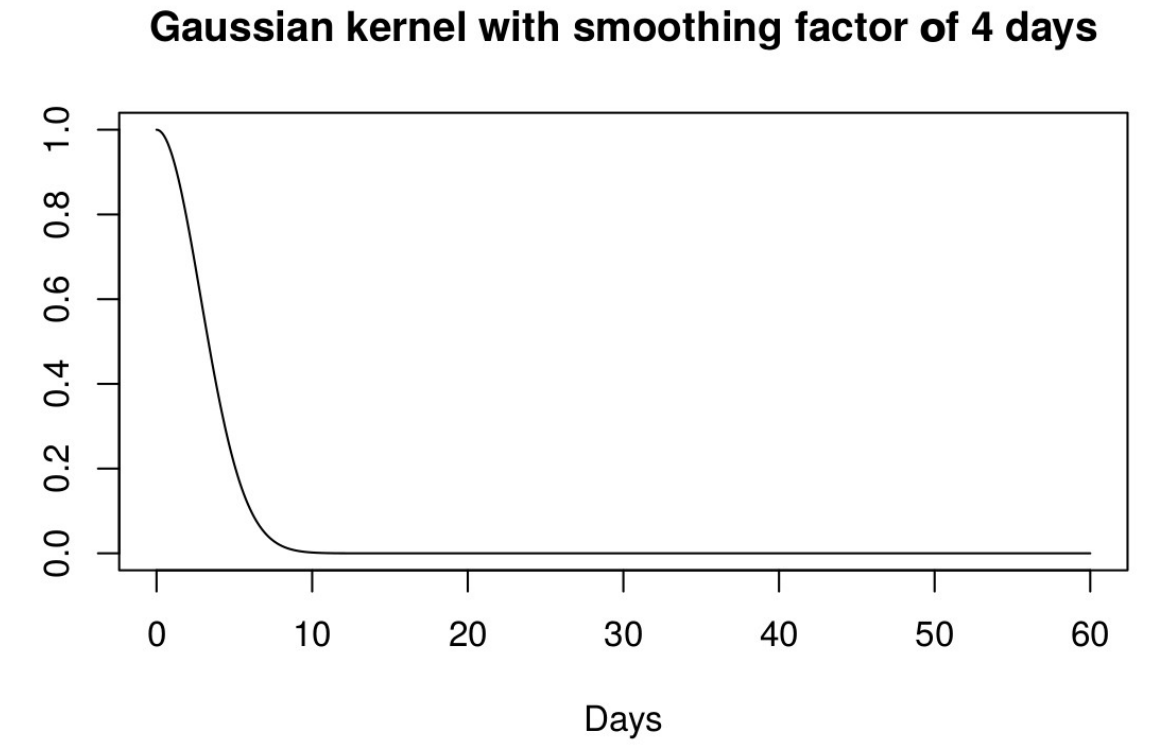
TDDE31-Lab 3

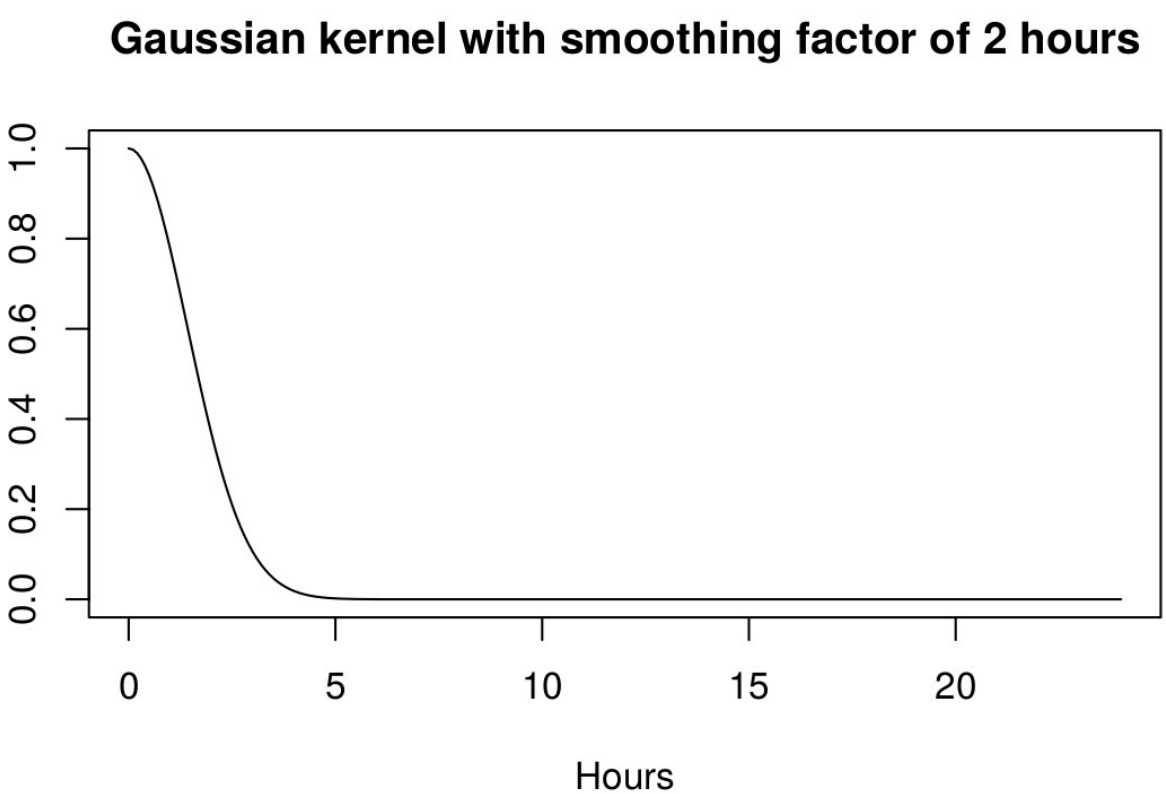
Kristian Sikiric (krisi211) & Pontus Svensson (ponsv690)

In this lab we implemented a kernel method to predict the hourly temperatures for a data and place in Sweden. The kernel is a sum of three Gaussian kernels; one that accounts for the distance from a station to the point of interest, one that accounts for the distance between the day of measure and the day of interest and one that accounts for the distance in time of the day between the measurement and the hour of interest.

Appropriate smoothing coefficients were chosen for the three kernels. $H_{distance}$ = 100 km, since the temperatures does not vary to much in a radius of 100 km in Sweden. H_{time} = 7200 seconds, since the temperature is roughly the same now as it was 2 hours ago. H_{date} = 4 days, since the average temperature is roughly the same in 4 day period. The following plots show the kernel width of the three kernels.




Gaussian kernel with smoothing factor of 4 days
0.0
0.2
0.4
0.6
0.8
1.0
0
10
20
30
40
50
60
Days


Gaussian kernel with smoothing factor of 2 hours
0.0
0.2
0.4
0.6
0.8
1.0
0
5
10
15
20
Hours

When summing the kernels, these are the following predictions we got for the date 2013-08-04:

('04:00:00', 15.065465074574973)

('06:00:00', 15.781374081558662)

('08:00:00', 16.390760279477167)

('10:00:00', 17.07640494368445)

('12:00:00', 17.75255829341874)

('14:00:00', 18.17981850520527)

('16:00:00', 18.49894262455248)

('18:00:00', 18.62653812427711)

('20:00:00', 18.29263113371574)

('22:00:00', 17.878096242599565)

('00:00:00', 15.89967687572718)

The predictions seems reasonable since it is supposed to be warm in the summer and it gets warmer during the middle of the day and then cooler again.

We see that the temperatures do not differ much from each other, this is because we use the sum of the kernels. This is probably because the kernels are independent from another, meaning that if a sation for example is 400 km away from the point of interest, this kernel will give a low value. But the date and time of the measurement might be close to the date and time we want to predict, so these kernels gives high values, and since we sum, the final kernel will get a high value even though the station is far away. This is fixed by multiplying the kernels.

These are the predictions for the same date when multiplying the kernels:

('04:00:00', 14.008705476306478)

('06:00:00', 16.72781738808453)

('08:00:00', 18.81847716354325)

('10:00:00', 20.48289489082506)

('12:00:00', 21.781137666303614)

('14:00:00', 21.81779624992602)

('16:00:00', 21.9619001089699)

('18:00:00', 20.301009745335254)

('20:00:00', 15.985560541775843)

('22:00:00', 14.335344839087762)

('00:00:00', 16.192445753046446)