ASSIGNMENT #3

Determining and removing drawbacks of exponential and running mean.

Group 2

Ekaterina Karmanova

Timur Chikichev

Iaroslav Okunevich

Nikita Mikhailovskiy

Part 1. Backward exponential smoothing

1. To conduct a simulation experiment, we have generated a true trajectory and its measurements by using random walk model:

$$X_i = X_{i-1} + w_i$$

Where: X – trajectory point,

w – normally distributed random noise

Size of trajectory: 300 points;

Initial condition: $X_1 = 10$;

Parameters of noise:

(Variance) $\sigma_w^2 = 28^2$;
(Mean) M = 0

To generate normally distributed noise in Python we were using function:

```
np.random.normal(mean, standart_deviation, size),
```

and definite iteration FOR.

By using the same method, we defined measurements:

$$z_i = X_i + \eta_i$$

With noise, that has variance:

$\sigma_n^2 = 97^2$.

2. We made a forward exponential smoothing with an optimal exponential smoothing coefficient. New coefficient α was found according to the new value of variances ratio.

$$\alpha = 0.25$$

Plot results of forward exponential smoothing:

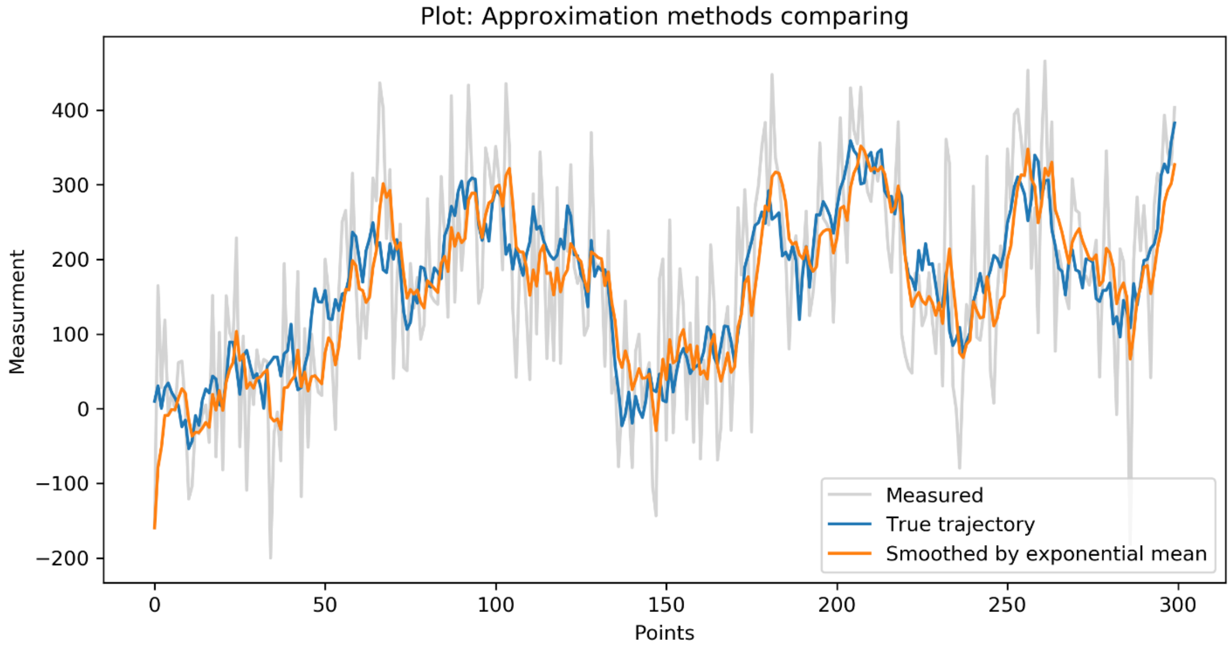

Figure 1 – Plot of result for 300 points (forward)

In these conditions results of exponential smoothing demonstrated significant shift (delay) of estimations.

3. To reduce that delay the backward exponential smoothing was applied to the previously smoothed curve.

Plot results of backward exponential smoothing to smoothed curve:

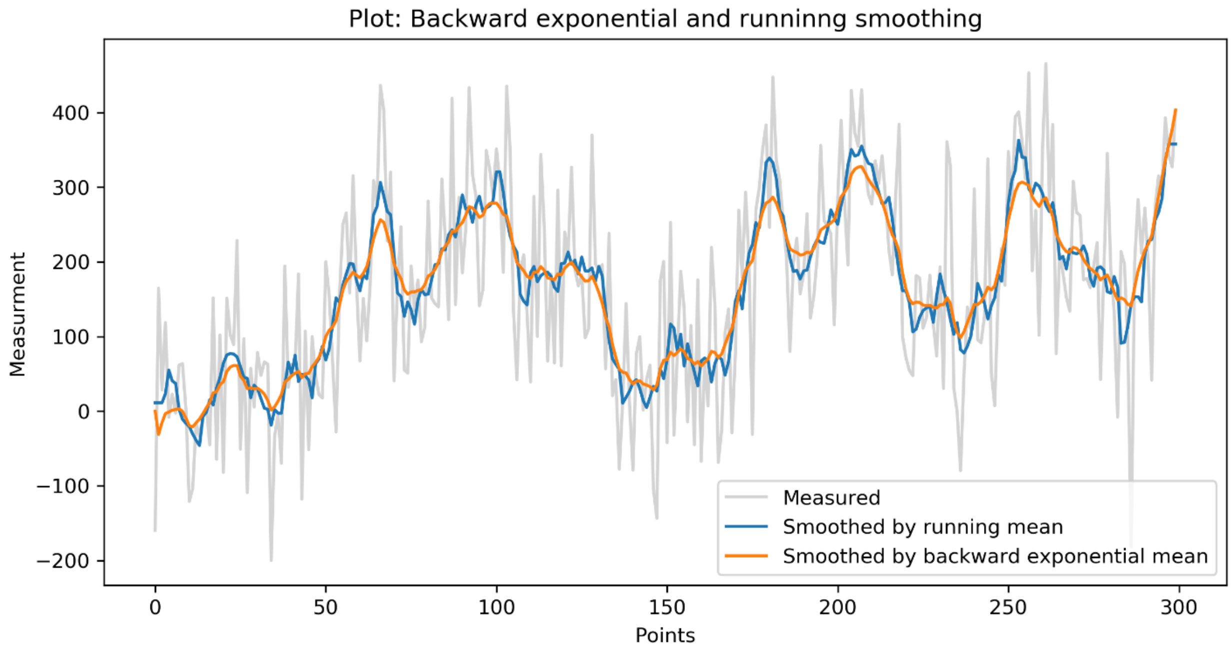

Figure 2 – Plot of result for 300 points (backward)

After backward smoothing of the curve the delay (shift) disappears. Running mean smoothing gives less fluent result than double exponential smoothing (forward and backward). The outcome is – backward exponential smoothing made data more smooth and easier to analyze, though it has some data lost.

Part 2. Drawbacks of running mean

First trajectory

1. We generated a process, which has insignificantly rate change and high measurement noise level.

   Parameters of trajectory:

$$X_i = X_{i-1} + V_{i-1}T + \frac{a_{i-1}T^2}{2}$$

$$V_i = V_{i-1} + a_{i-1}T$$

Size of trajectory is 300 points;

Initial condition was $X_1 = 5$; $V_1 = 0$; $T = 0.1$;

Variance of noise: $\sigma_a^2 = 10$;

Parameters of measurements:

$$z_i = X_i + \eta_i$$

Variance of noise: $\sigma_n^2 = 500$

Plot results for the trajectory and velocity:

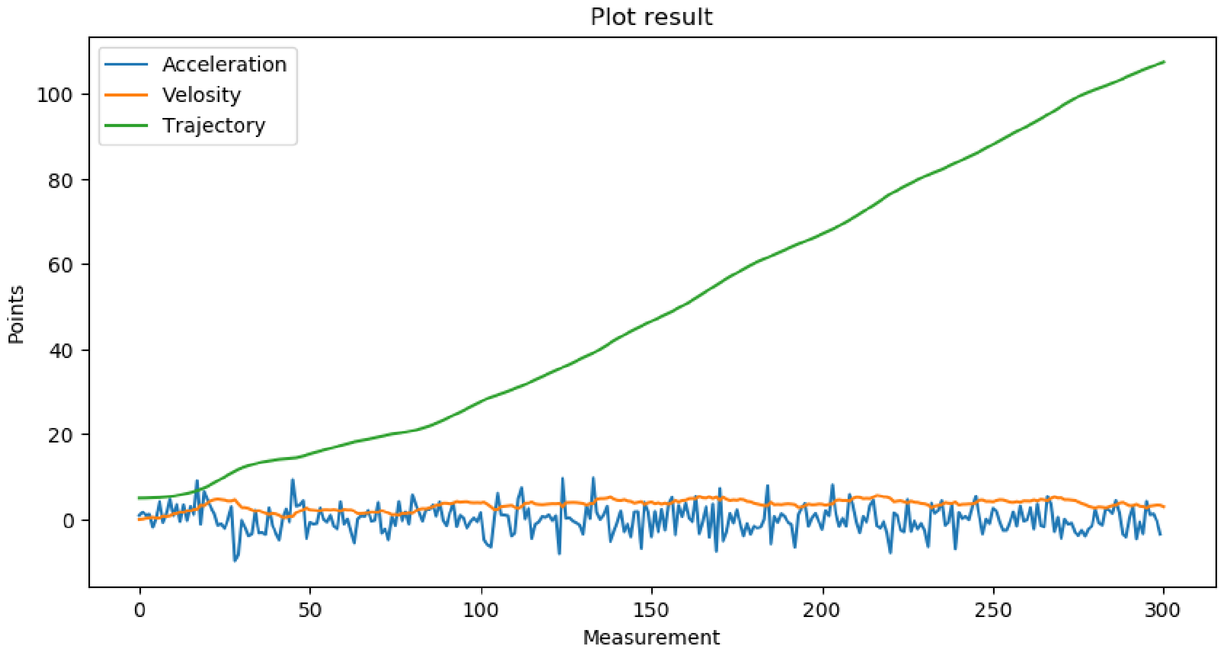

Figure 3 - Plot of result for the first process

Plot results for the trajectory and measurements:

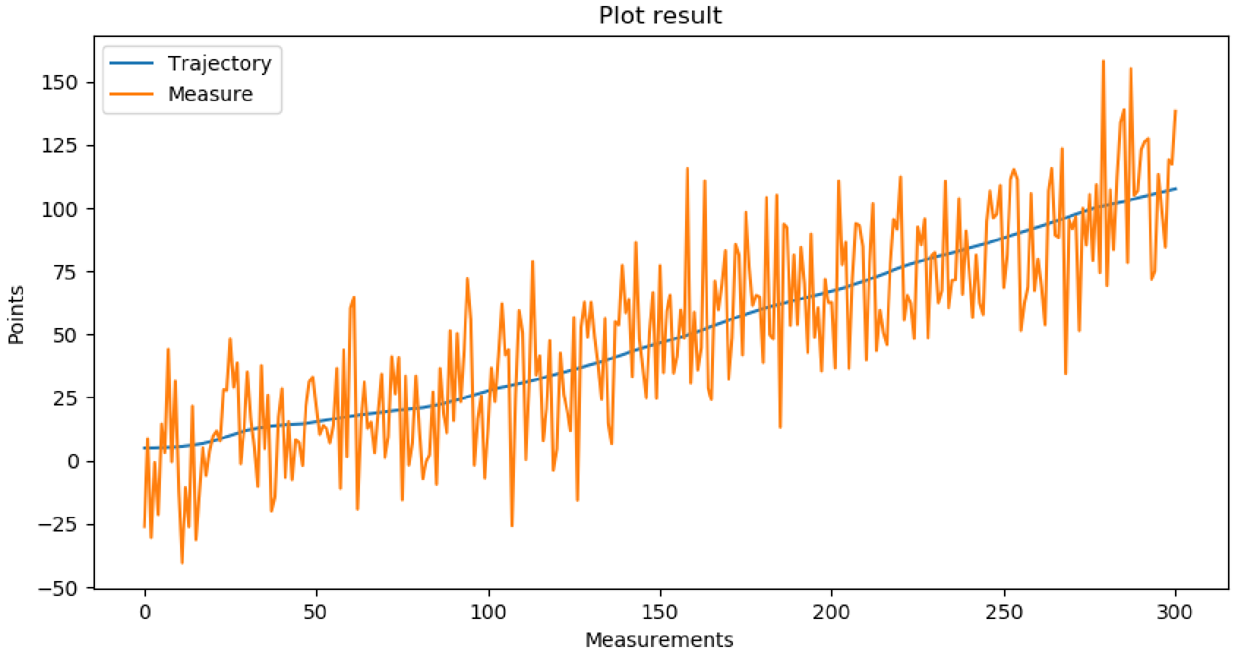

Figure 4 - Plot of result of measurements for the first process

2. To determine the window size of running mean we have to analyze the trajectory shape and level of noise in measurements. According to the fact that the trajectory is close to line and the noise is frequent and great for more effective filtration of measurement errors it is better to choose a wide window (so we did not test M<20). To make the process of finding the best window size more reasonable we decided to use sequential approximation method (the least step was 13 points) and calculate mathematical expectations for different values.

The best result for window size:

M =40

Sum_indicators = 2998

Plot results for different window width values, including the best option:

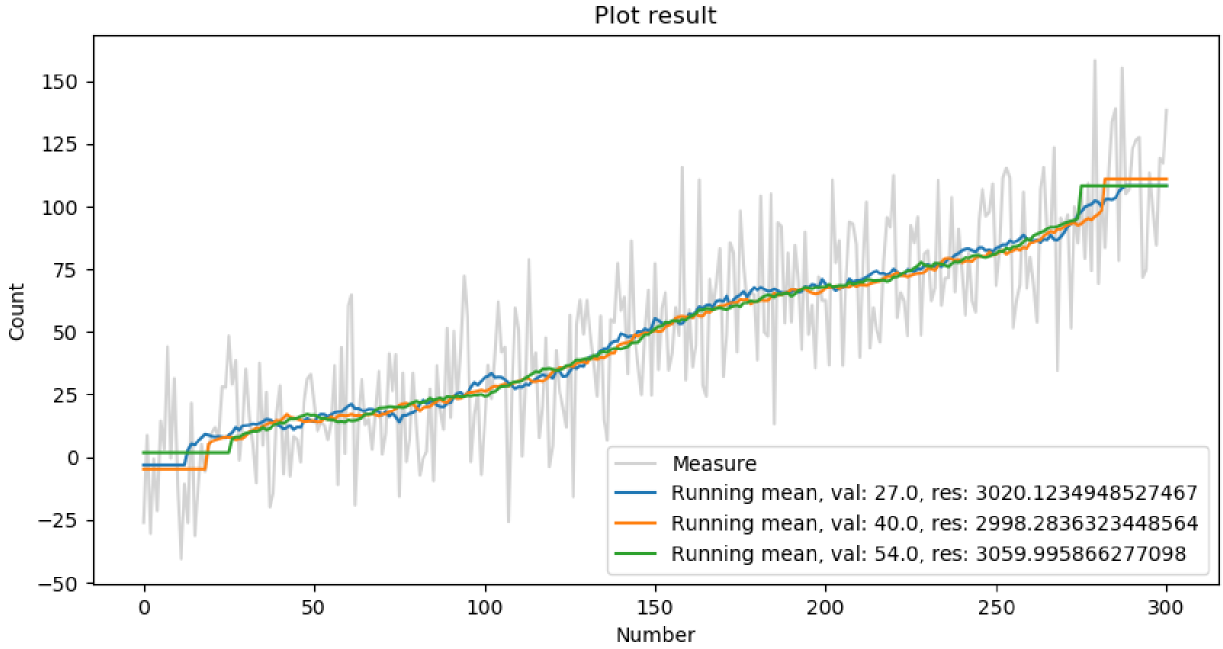

Figure 5 – Plot of result for 300 points for different window width (top 3 best results)

3. To find exponential coefficient (alfa) we chose several values and distinguish the best option according to calculated indicators for each option. The lower the indicators are, the better smoothing effect is achieved.

The best result for window size:

$$\alpha = 0.36$$

Sum_indicators = 2478

Plot results for different exponential coefficient values, including the best option:

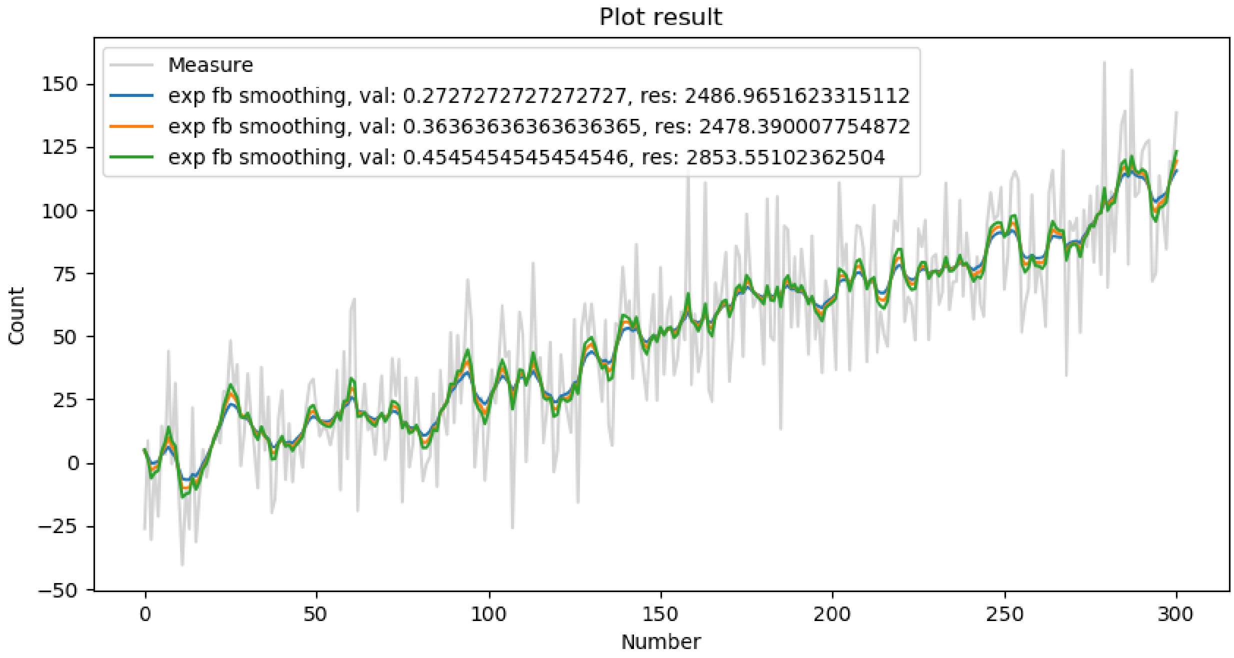

Figure 6 – Plot of result for 300 points for different α value (top 3 best results)

According to the outcomes which have been achieved (the values and plots), it seems that exponential smoothing method is more efficient for this process, because the sum of deviation and variability indicators is less. From the plots 5 and 6 we can see that the shift (delay) in exponential smoothed curve is less and, though the sensitivity is quite high, the resulting sum of indicators is lower than with the running mean method.

Second trajectory

4. The next trajectory is periodical. However, the initial steps are the same as for the previous task (first trajectory). We generated a cycle process, which has insignificantly measurement noise level.

Parameters of trajectory:

$$X_i = A_i \cdot \sin(\omega i + 3)$$
$$A_i = A_{i-1} + w_i$$

Size of trajectory is 200 points;

Period of oscillations is T=32 steps;

Initial condition was $A_1 = 1$;

Variance of noise: $\sigma_w^2 = 0.08^2$;

Parameters of measurements:

$$z_i = X_i + \eta_i$$

Variance of noise: $\sigma_n^2 = 0.05$.

5. Firstly, the window size for running mean smoothing was detected: M=13.

Plot results for RM smoothed curve with window size 13 points:

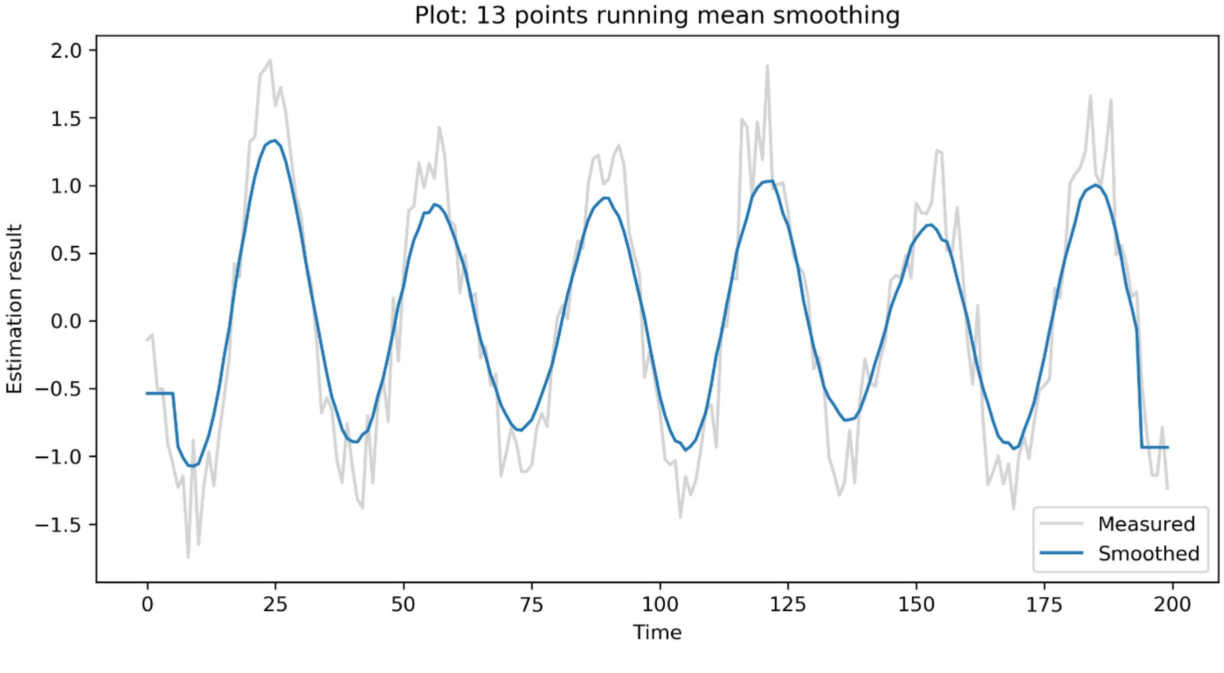

Figure 7 – Plot of result for 200 points with window size 13 points

We see that the period of oscillation of smoothed curve is equal to the initial trajectory and the amplitude is nearly the same.

6. Secondly, the window size has been changed: M_new=17.

a) To produce inverse oscillation by running mean smoothing, we should include at least 3/4T points from both sides. Middle points have their flipped copies that have opposite signs, that is why they compensate each other and that is the reason why do other opposite sign neighboring points are included in the equation. That means for every point Z_i in running mean smoothing equation points ($Z_{i-3T/4} - Z_{i+3T/4}$) should be included and the ratio between the window size and the oscillation period is:

$$M = 1.5 \cdot T \rightarrow T = \frac{M}{1.5}$$

Plot results for RM inversed smoothed curve with window size 17 points:

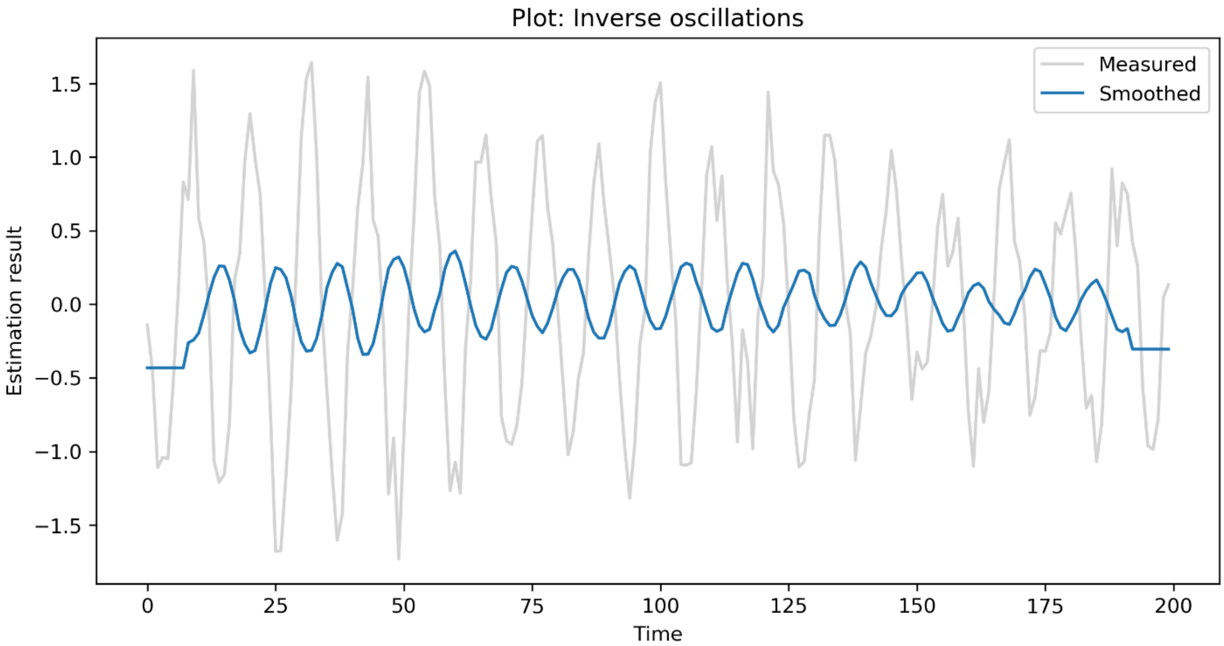

Figure 8 – Plot of result for inversed result with window size 17 points

According to the figure the amplitude is lower than the measurements have, despite the fact that the greatest number of points were used. If we go further from that window ($Z_{i-3T/4} – Z_{i+3T/4}$) values of smoothed points will just decrease because more points will start compensate each other. This is the disadvantage of running mean method implemented to cycle process.

b) To avoid oscillation in smoothed signal we just need to take the window size for the running mean equal to the period of oscillation of true signal, so that the opposite points of trajectory will compensate each other:

$$T = M = 17 \ steps.$$

Plot results for RM smoothed curve with zero oscillation with window size 17 points:

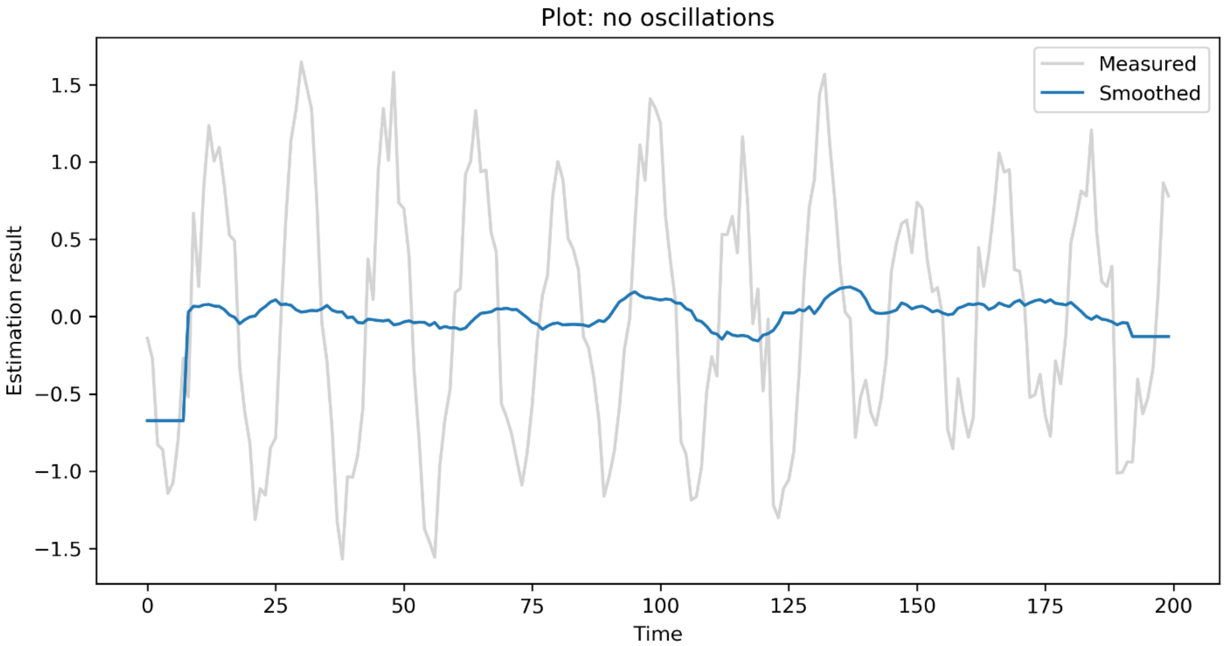

Figure 9 – Plot of result for zero oscillation result with window size 17 points

The real smoothed curve is pretty close to a line. Some waves are caused by asymmetry of the measurement cycle curve.

c) The last task is to change insignificantly the oscillations. To achieve this result points without flipped copies should be taken. The less flipped copies are in running mean equation, the closer to initial value the result is:

$$M \leq 0.5 \cdot T \rightarrow T \geq 2 \cdot M.$$

Plot results for RM smoothed curve with oscillation close to initial with window size 17 points:

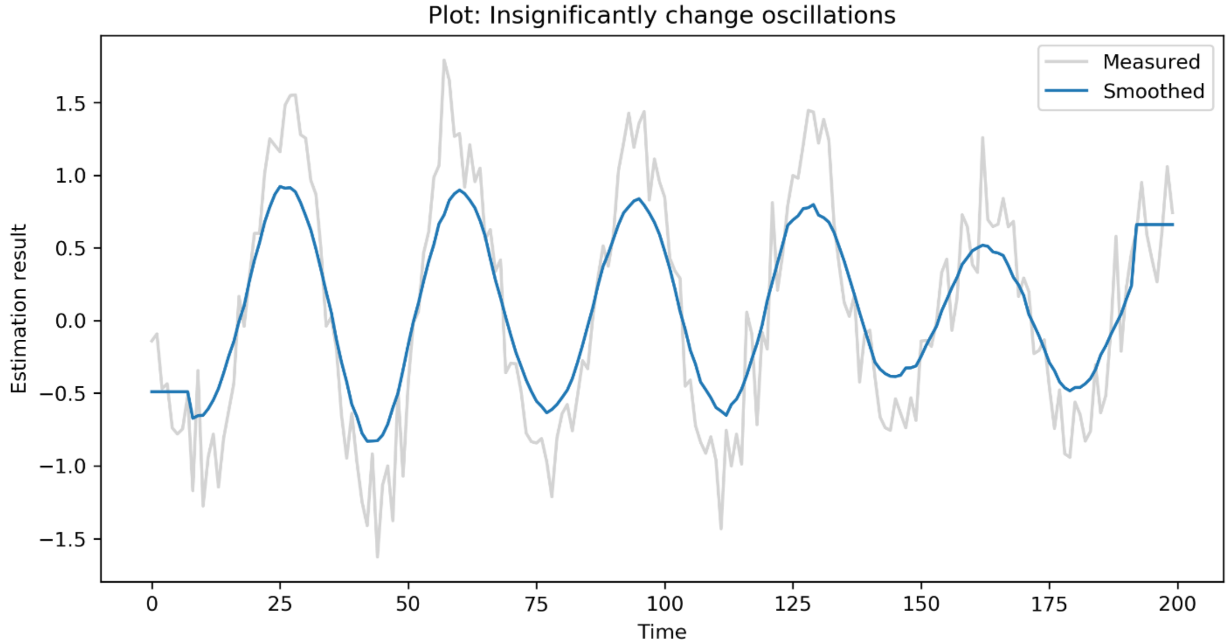

Figure 10 – Plot of result for close oscillation result with window size 17 points

The smoothed curve has nearly steady amplitude very close to initial one. The period of oscillation is equal to the initial.

Conclusions

During this Assignment we compared backward exponential and running mean smoothing methods: advantages and disadvantages of applying each method.

We learnt how to determine optimal smoothing coefficient in exponential smoothing and window size empirically (with less effort). Moreover, we found out that determination of window size for running mean didn’t depend on exponential smoothing coefficient for the known initial trajectory.

We also obtain new skills of how to transform measurements of cycle process via smoothing into inverse or non-oscillating curve – which can be useful for further analyses of received data.