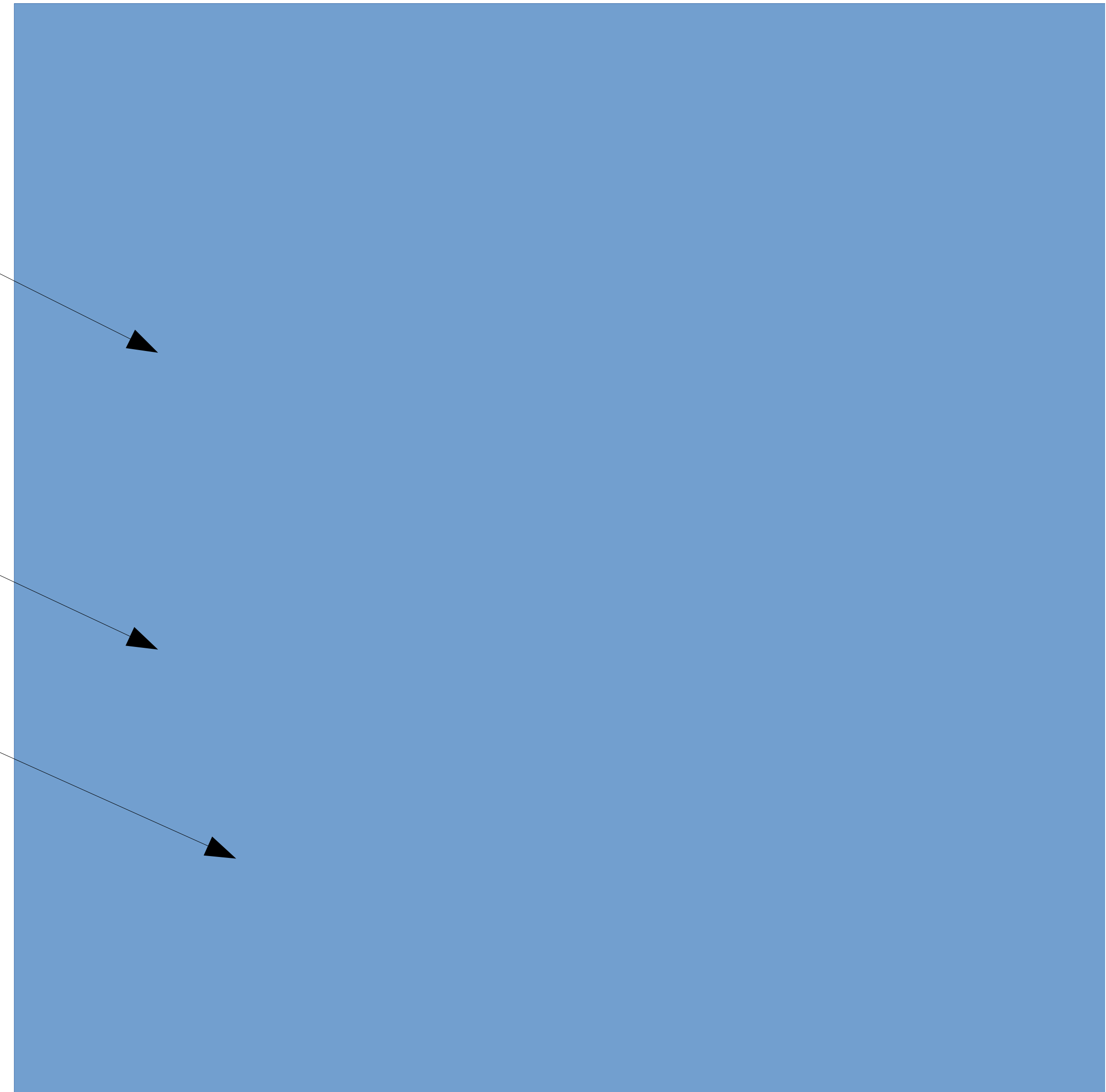
(Time) series

- Sequence of data points over some parameter
- Analysis methods can be applied to any data with some natural ordering
- Main topic: Separation of signal components
- Main goal: Predicting the future (model)

Components

- Seasonal
- Trend
- Cyclic
- Irregular (noise)



Classical decomposition (~1920s)

- Additive: $y = T + S + C + I$
- Multiplicative: $y = T \cdot S \cdot C \cdot I$
- Recipe:
  - Find trend T and detrend: $y' = y - T$ $(y' = y/T)$
  - Find seasonal comp. …: $y'' = y' - S$ $(y'' = y'/S)$
  - Remainder consists of cyclic and irregular comp.
- We will focus on:
  - 2 detrending methods (moving avg. & fourier ana.)
  - A method for finding seasonal comp. (fourier ana.)
  - A method for finding cyclic comp. (autocorr.)

Detrending - Moving average

Average over a moving window of data

$$\hat{T}_t = \frac{1}{m} \sum_{j=-k}^{k} y_{t+j}$$

- Add weights if necessary:
  - Gaussian
  - Exponential
  - Triangle
  - Whatever you can think of
- Pro: Easy and robust
- Con: Will usually incorporate part of other comp. (median?)

Slime mould experiment

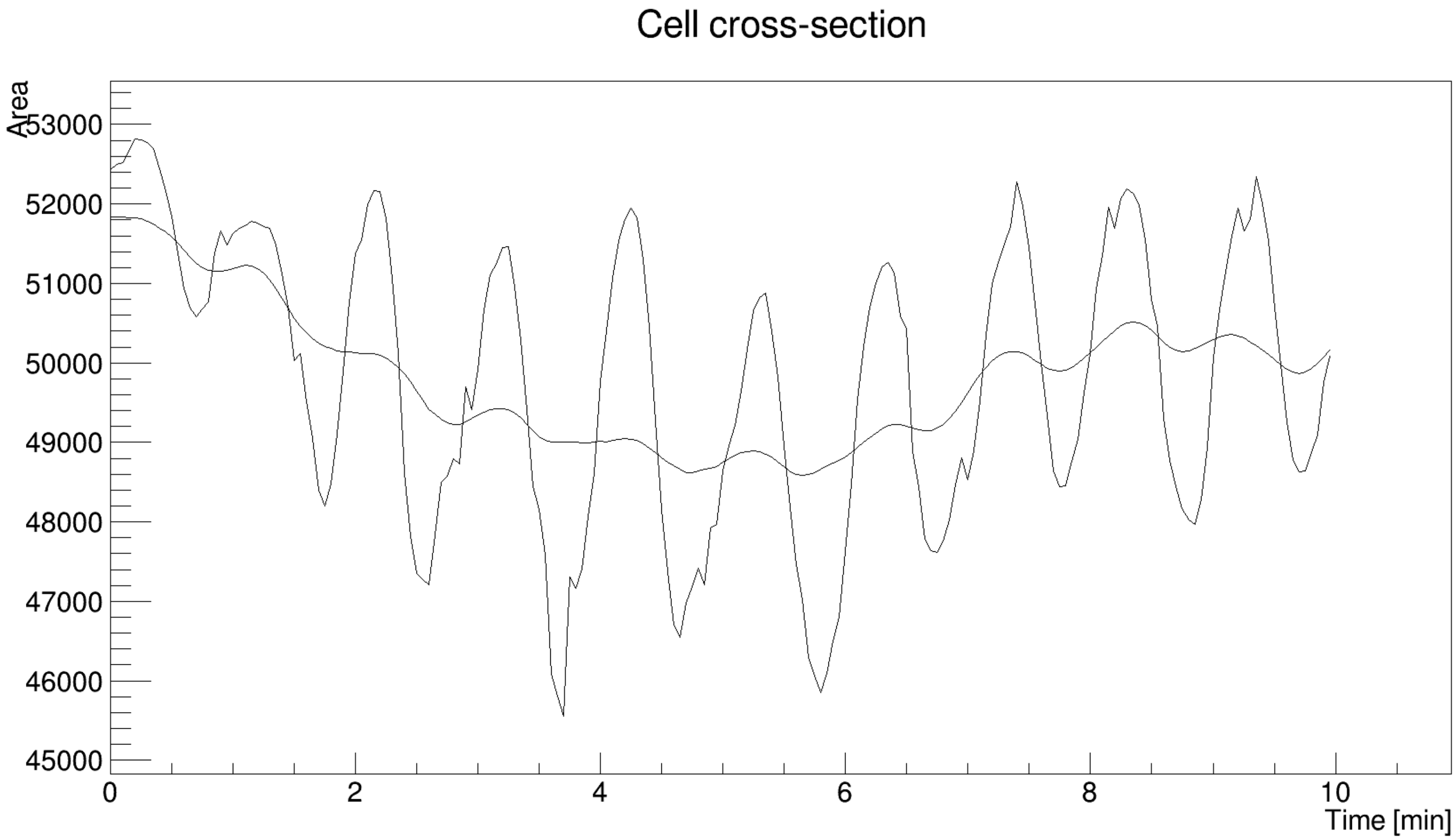
- Show: https://www.youtube.com/watch?v=GY_uMH8Xpy0
- Mention visible “vibrations” (~scale minute[s])
- What is a slime mould:
  - One cell, move stuff around inside
  - Contractions (almost periodic), pumping function
  - Microplasmodium does the same
- Preparations (and data aquisition):
  - Remove slime in centrifuge
  - Resuspend in growth medium on petri dish
  - Remove liquid after spreading
  - Put under light microscope
  - Wait and take pictures
- Analysis:
  - Do a lot of image processing to get a binary picture
  - Analyze boundary oscillations and analyze area oscillation

Exercise

- Download 3 files:
  - images.json.gz – data (explain data origin here!)
  - timeSeries.py – main program
  - timeSeries_functions.py – useful functions
- You will only write very short pieces of code
  - Don't write ROOT code (use DrawGraph, saves time)
  - Don't write your own functions (only fill in missing code)
  - Carefully read the comments on each function (save yourself a lot of time)
- All steps are explained in detail in the main program

Exercise

- Work on the exercise until you have a moving average of the cell area:



Fast fourier transform (FFT)

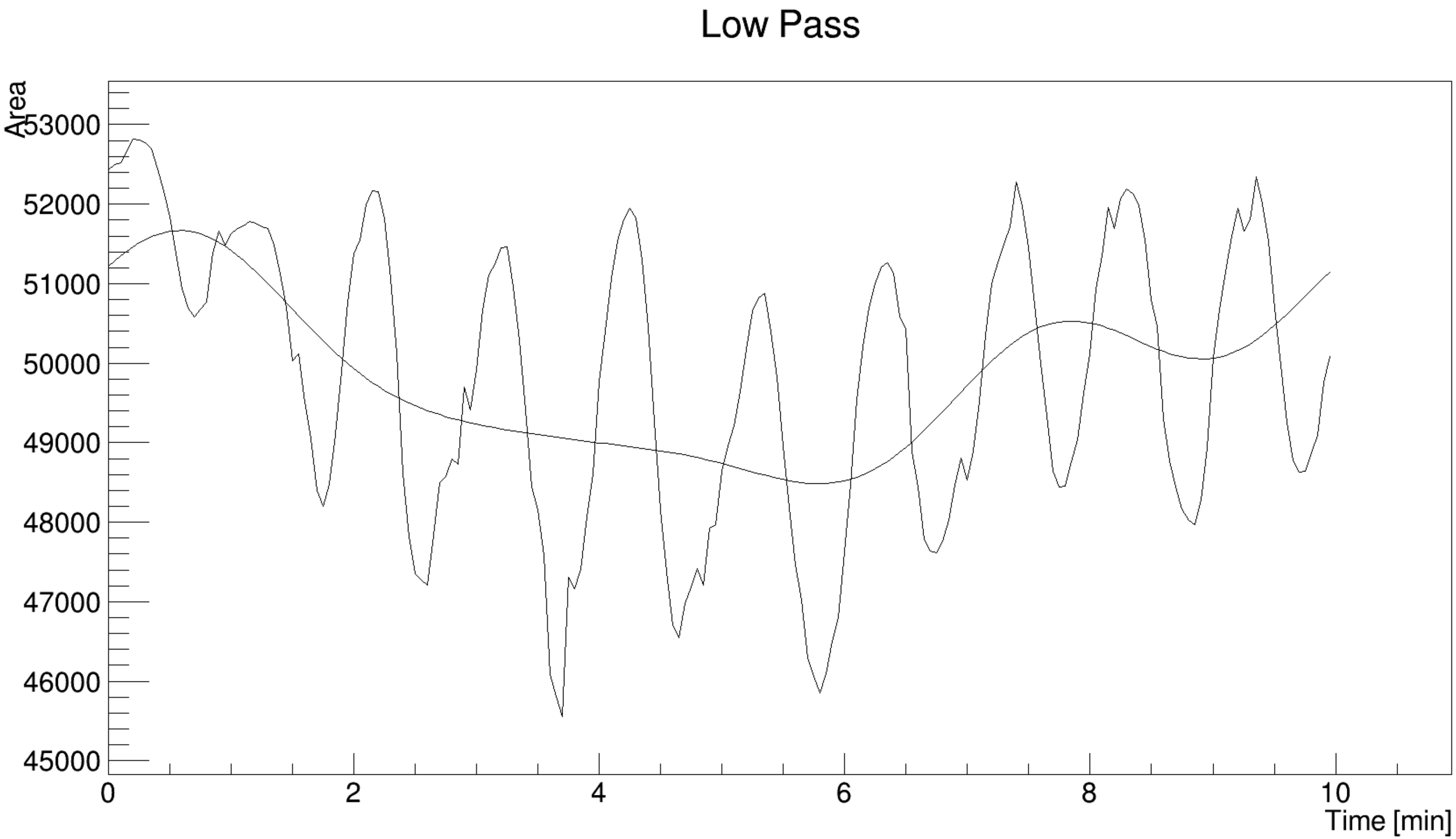
- Algorithm for DFT
  - Fast! (scales linearly)
- Requires same distance
  - Interpolate if necessary
- Returns spectrum of complex numbers over frequency
  - Represents amplitude & phase in one number
- I am assuming fourier transforms are familiar
  - Tell me if I am wrong
  - Do you have any questions regarding this (technical stuff?)

Uses of FFT

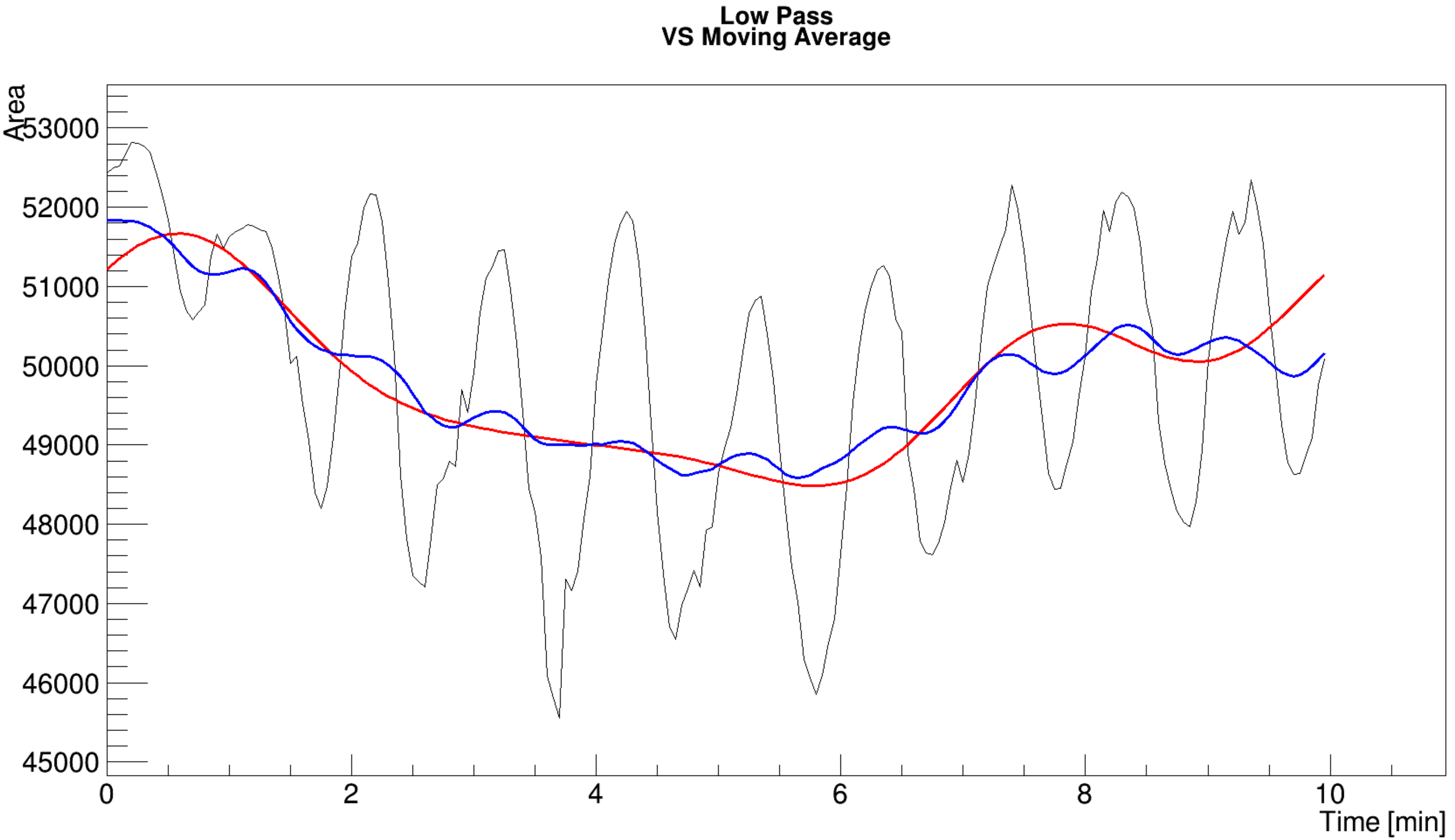
- Transform; modify spectrum; transform back
- Compression
  - Identify important frequencies, delete the rest
- Detrending (similar to compression)
  - Separate into high and low frequencies
    - Low frequencies are trend (Low-pass filter)
    - High frequencies are detrended data (High-pass filter)
- Seasonal components are peaks frequency domain
  - Separate seasonal component by isolating peaks
    - Peaks are seasonal component
    - Rest is deseasonalized data
- Carefull: Not usefull for cyclic components
  - Reason: Phase shift between cycles

Exercise

- Work on the exercise until you have applied a low pass filter to the cell area



What do you think?



Cyclic (& irregular) component

- Any correlation without exact timing
  - Oscillations with phase shifts
  - Feedback of past data on current data
    - Flights (Peak of return flights after peak in outgoing ones)
    - Random walk (maximum step size, questionable if cyclic)
    - Influenza (higher chance for epidemics after quiet times)
- Hard to model
  - Correct model depends on exact mechanism

Autocorrelation

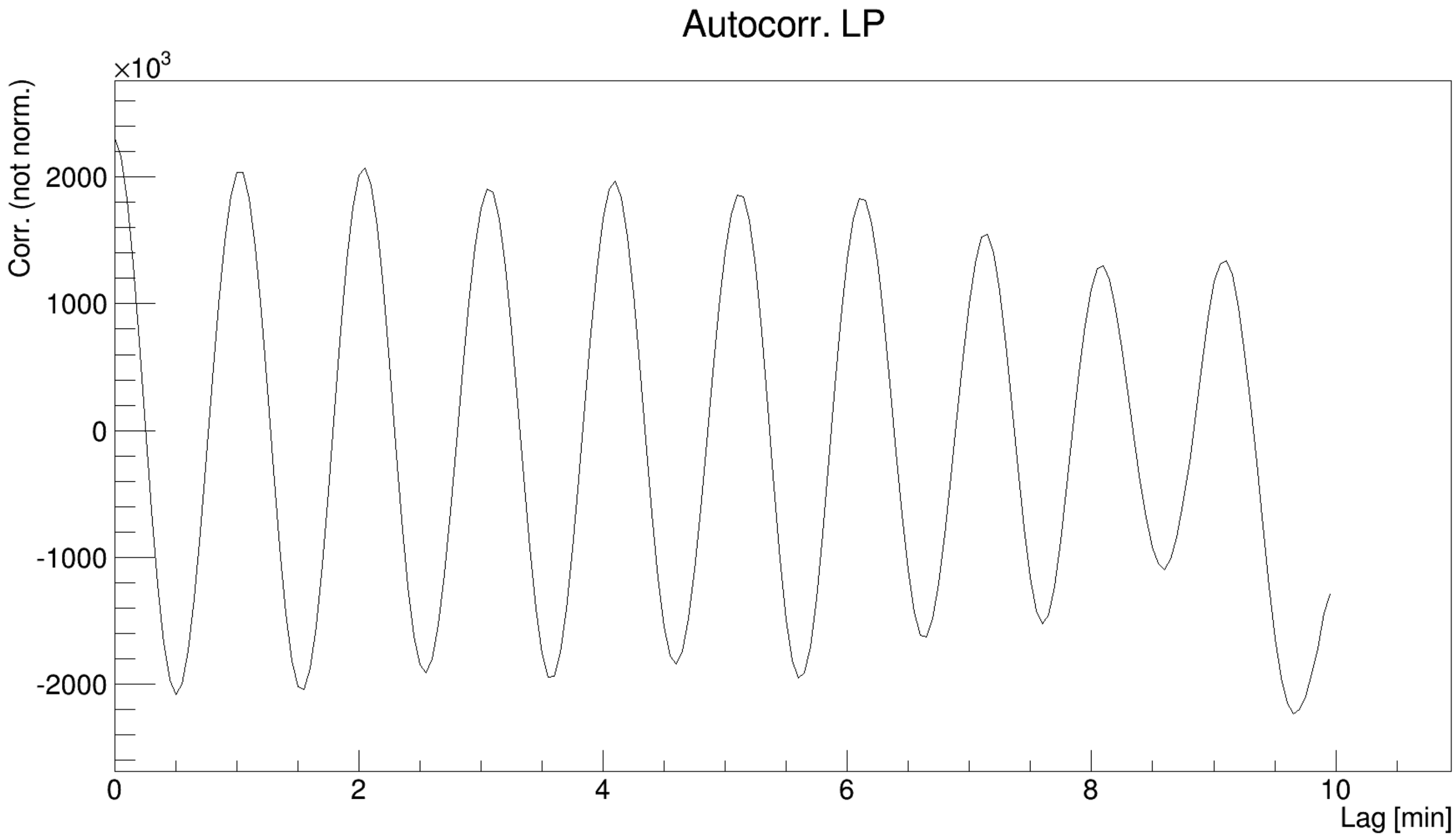
- Relationship between observations of the same variable at different times:

$$R(\tau)=\frac{\frac{1}{N-\tau}\sum_{t=1}^{N-\tau}(X_t-\mu)(X_{t+\tau}-\mu)}{\sigma^2}$$

- Confidence intervall: $CI=\frac{\pm 2}{\sqrt{N-\tau}}$

Exercise

- Continue with the exercise until you computed the autocorrelation and the contraction frequency



Autoregressive model

- Suppose you found significant autocorrelation
- Assume random process with linear dependence on last few timesteps

$$X_t = c + \sum_{i=1}^{p} \varphi_i X_{t-i} + \epsilon_t$$

$$\varphi_1, \ldots, \varphi_p : \text{model parameters}$$
$$\epsilon_t : \text{noise term}$$

- Fit using the appropriate method for your noise term (least-squares for gaussian)
- Now that you have a model, you can use it to extrapolate (while propagating errors in each step)
- Combine this with other components by fitting their residuals