IL2233

Lab 1: Time Series Visualization and Feature Extraction

Shawn Nagar, Leon Moll

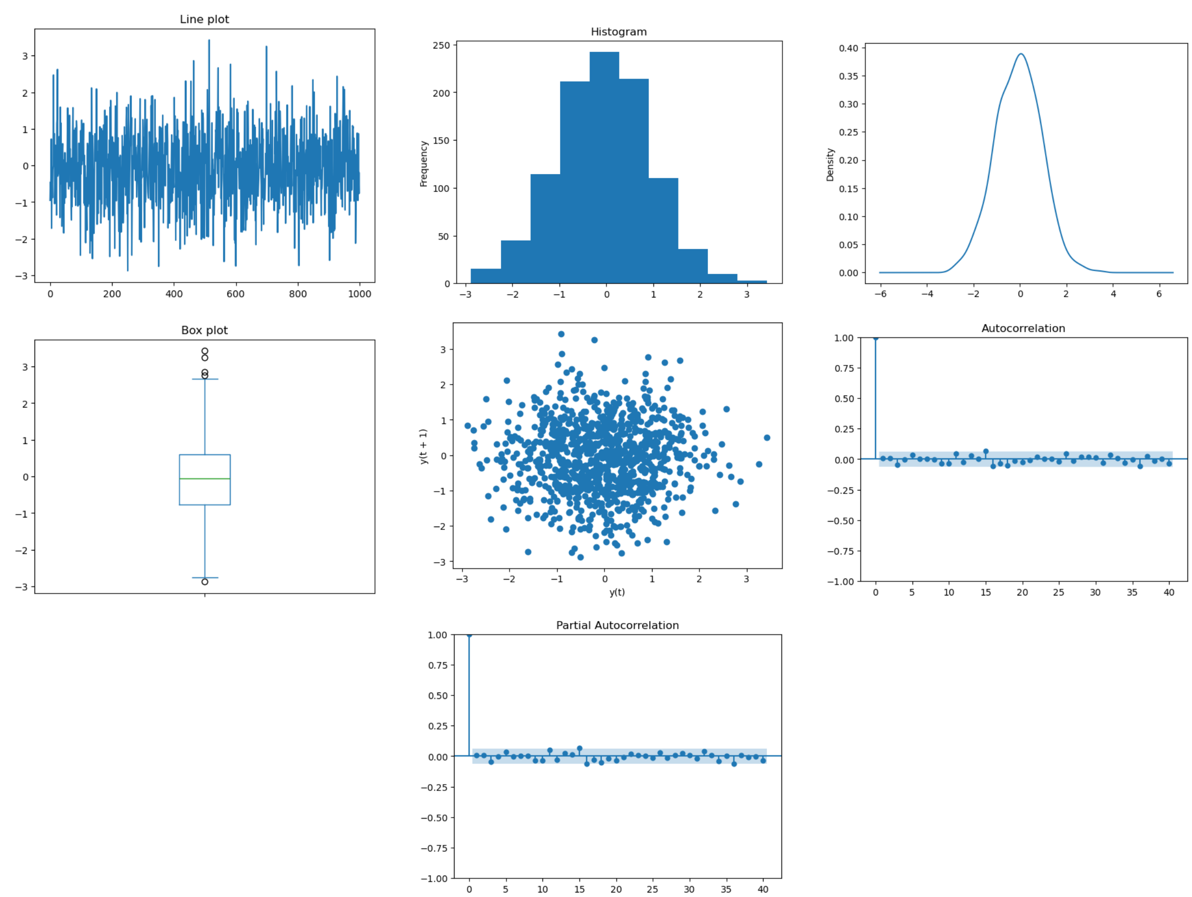
April 9, 2024

1 Exploratory Data Analysis

1.1 White noise series

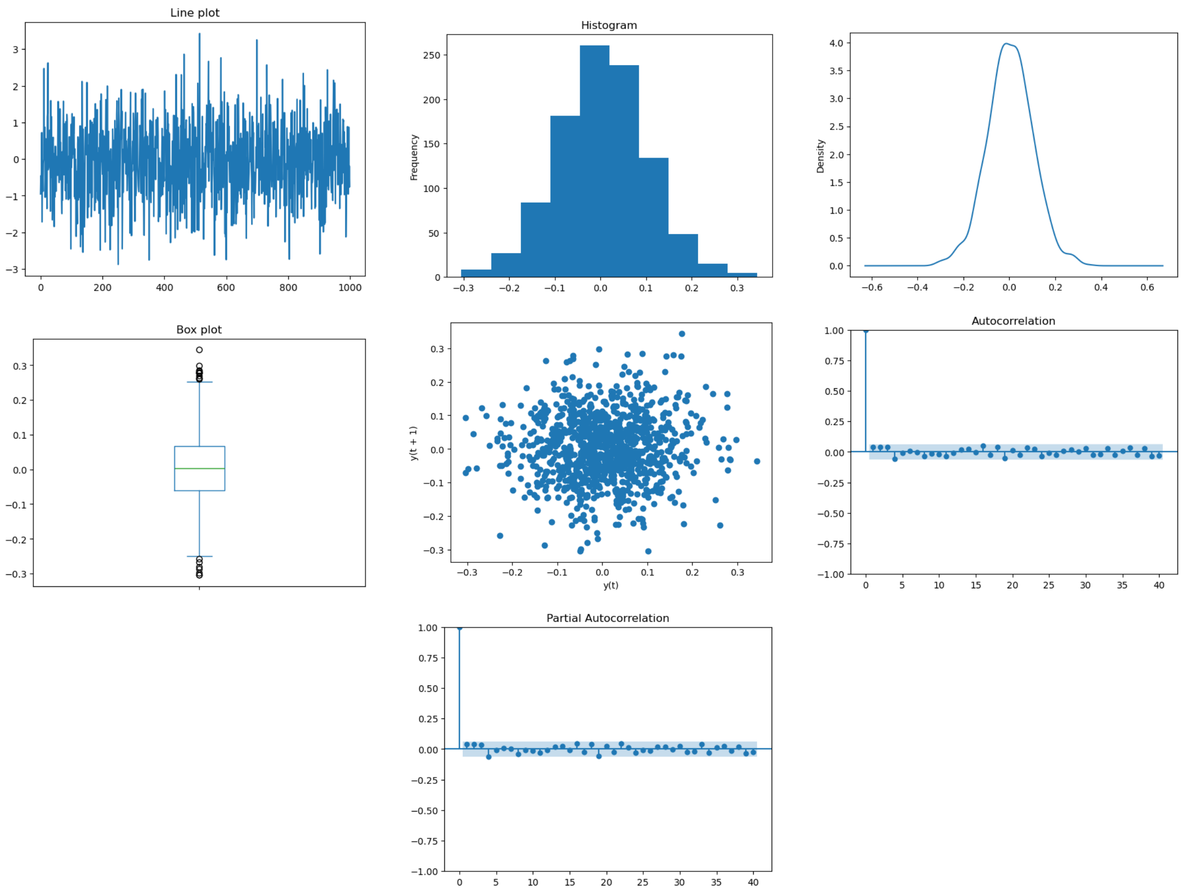
We generated the white noise and average value white noise series. Below are the plots and values for both series.

1.1.1 Single white noise series



Mean	Standard Deviation
-0.07120429237582035	0.9922827480232803

1.1.2 Combined average white noise series



Mean	Standard Deviation
0.004170518169858113	0.09882040818504637

1.1.3 Ljung-Box test

Statistic	p-value
6.391248	0.781391

The p-value is > 0.05, therefore we accept the Null hypothesis. Thus the series is random.

1.1.4 Augmented Dickey- Fuller test

The p-value is < 0.05, therefore we reject the Null hypothesis. Thus the series is stationary.

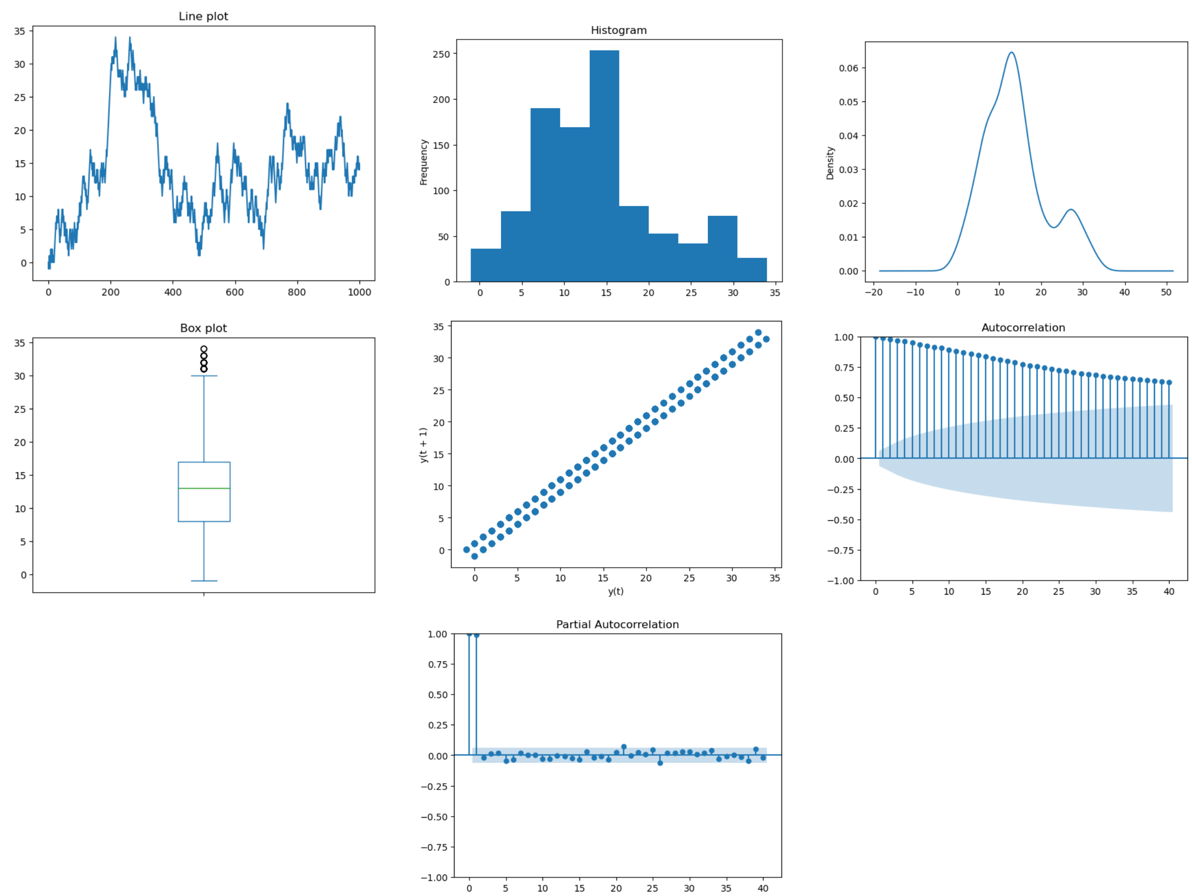
Statistic	p-value
-31.336703	0.000000

1.2 Random-Walk Series

1.2.1 Realizing a random-walk series

We created a series where the next value depends on the previous value.

1.2.2 Plotting the results



1.2.3 Ljung-Box test

Statistic	p-value
8942.124393	0.0

The p-value is < 0.05, therefore we reject the Null hypothesis. Thus the series is not random.

1.2.4 Augmented Dickey- Fuller test

Statistic	p-value
-2.519923	0.110702

The p-value is > 0.05, therefore we accept the Null hypothesis. Thus the series is not stationary. We can use differencing to generate a one order difference which would then be stationary.

1.2.5 Questions

What methods can be used to check if a series is random? Describe both visualization and statistic test methods. For visualization methods, we can use the different graphs:

- line plot: check if we can see a continuous line
- histogramm/density: check if we see a guassian distribution
- box plot: generate a series of box plots and check if there is no pattern
- lag plot: check if we have different clusters
- (partial) autocorrelation: check if there is no correlation with other values

For statistic methods we use the Ljung-Box test. If the p-value is > 0.05 the series is random.

What methods can be used to check if a series is stationary? Describe both visualization and statistic test methods.
For visualization methods, we can use the different graphs:

- line plot: check if we do not see a trend
- histogramm/density: a guassian distribution can be an indication for a stationary series
- box plot: generate a series of box plots and check if there is no trend
- autocorrelation: a slowly decaying autocorrelation plot is an indication for a stationary series

For statistic methods we use the ADF test. If the p-value is < 0.05 the series is random.

Why is white noise important for time-series prediction?
The first step of EDA should be to check if the signal is just white noise. Therefore we have to be familiar with it. Another point is that that series of errors of a time series forecast model should be white noise.

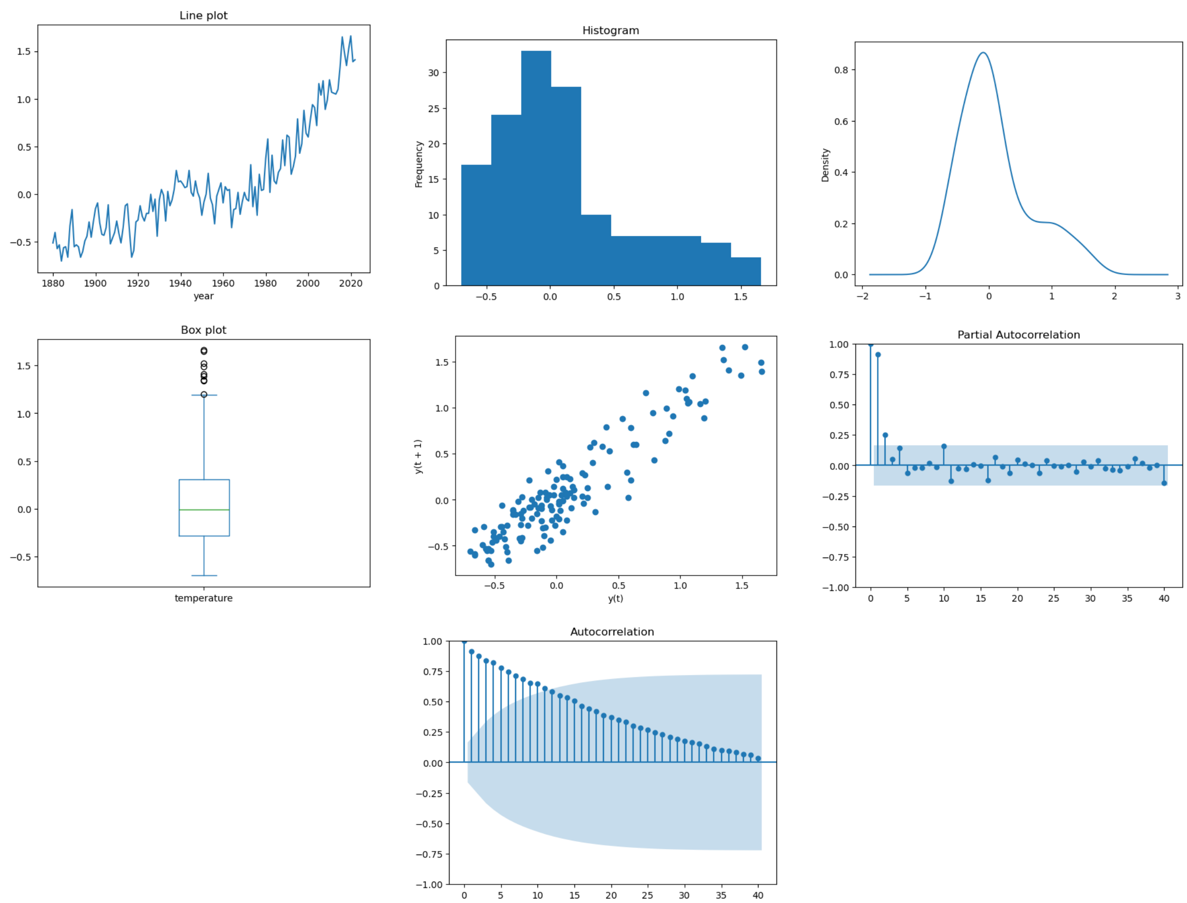
What is the difference between a white noise series and a random walk series?
White noise is stationary while random walk series is not, as a value of a random walk series

always depends on the previous value.

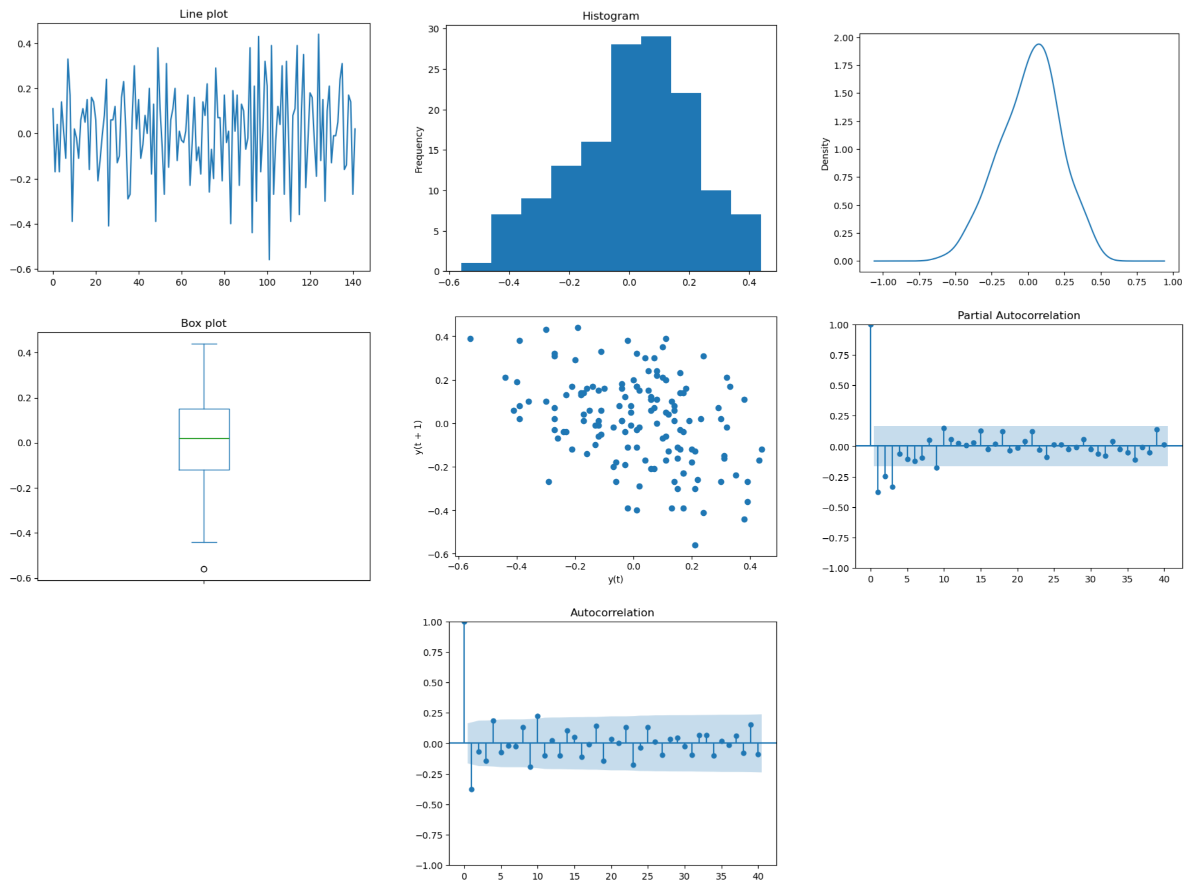
Is it possible to change a random walk series into a series without correlation across its values ? If so, how? Explain also why it can.
This can be done by first order differencing. With this we remove the value of the predecessor from the current value. This way only the stationary part is left.

1.3 Global land and temperature anomalies series

1.3.1 Plotting of anomalies series



1.3.2 Plotting of first order difference of anomalies series



1.3.3 Test if the original and the differenced temperature anomaly series are random or not.

To test randomness we did the Ljung-Box test.

	lb p-value
Original	4.694010e-185
Differenced	8.670322e-07

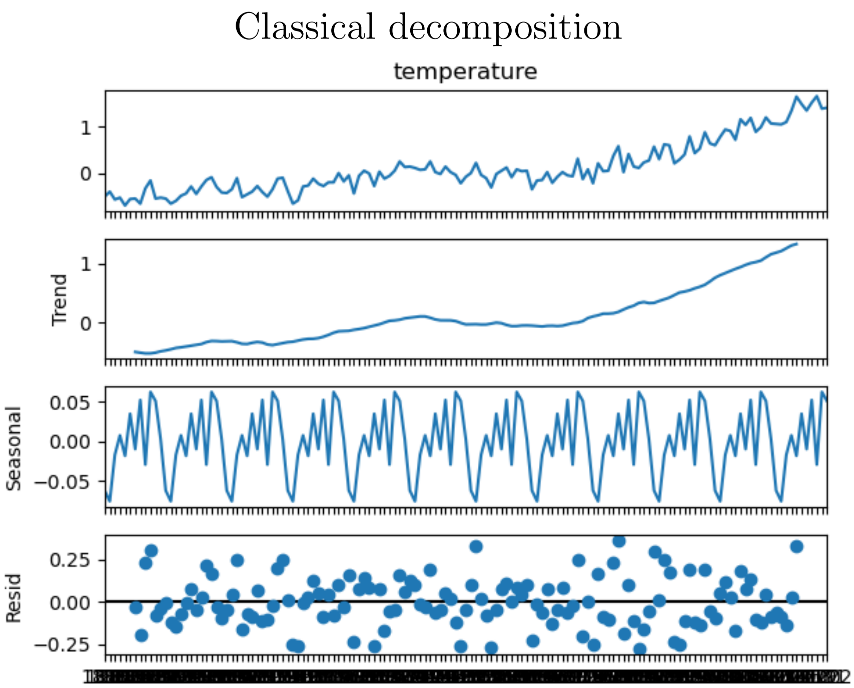
The results show that both series are not random

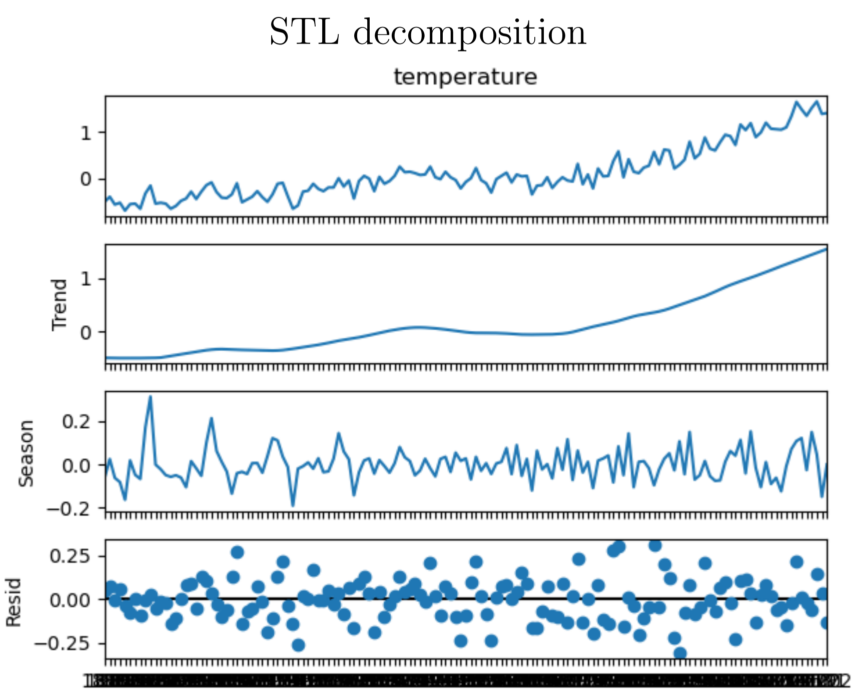
1.3.4 Test if the original and the differenced temperature anomaly series are stationary or not.

To test stationarity we did the ADF test. The results show that the differenced series is stationary while the original series is not.

	p-value
Original	1.000000
Differenced	0.000000

1.3.5 Classical and STL decomposition





1.3.6 Questions

What is a stationary time series?
A stationary time series is a series which properties (e.g. mean, variance) are independent of time.

If a series is not stationary, is it possible to transform it into a stationary one? If so, give one technique to do it.
This can be done by differencing.

Is the global land temperature anomaly series stationary? Why or why not?
This series is not stationary, as the p-value of the ADF test is > 0.05. It is also visible in different graphs: From the Line plot we can say that there is probably an upwards trend and in both decompositions this shows clearly in the "Trend" section.

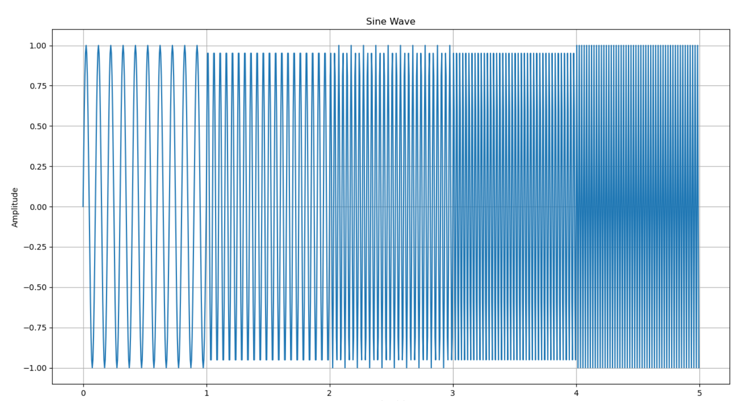
Is the data set after the first-order difference stationary?
This series is stationary, as the p-value of the ADF test is < 0.05.

Why is it useful to decompose a time series into a few components? What are the typical components in a time-series decomposition?
Decomposing helps to better understand the time series, as it allows us to analyze the different components individually. The typical components in a time series are trend-cycle, seasonal and remainder.

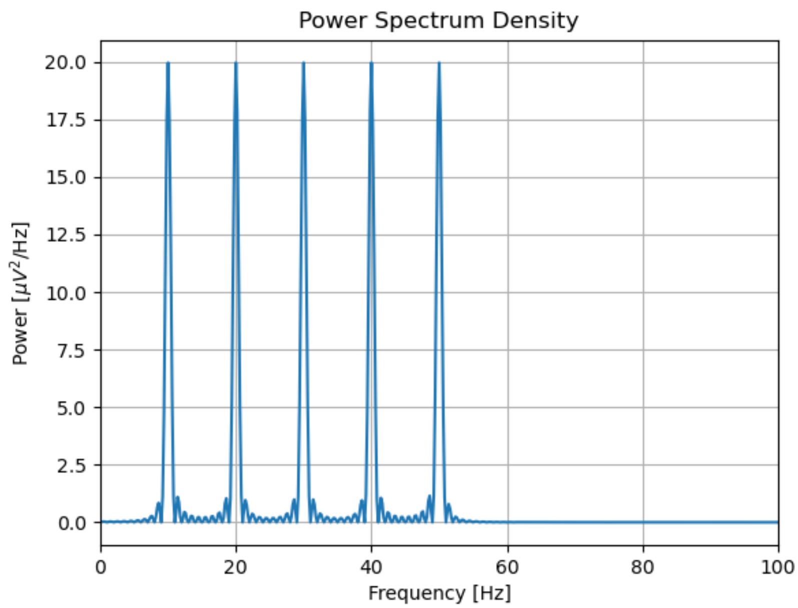
2 Feature Extraction

2.1 Frequency components of a synthetic time-series signal

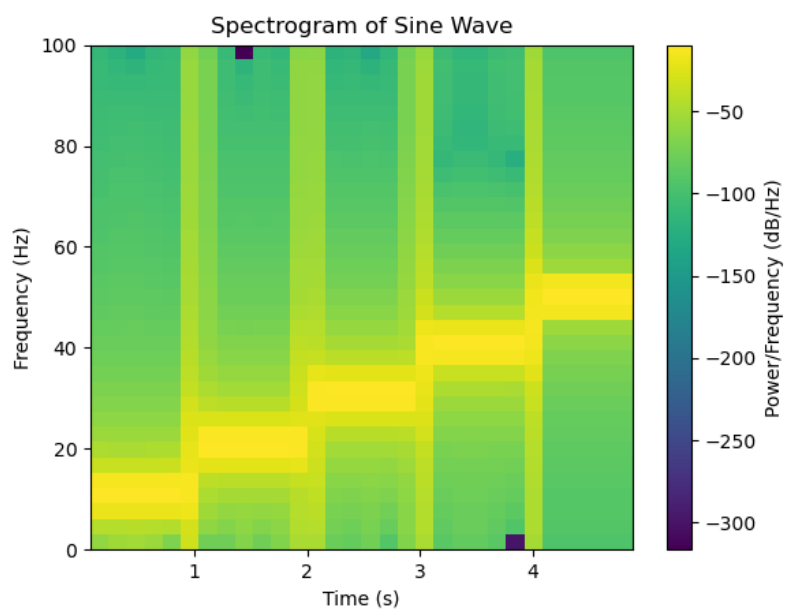
2.1.1 Line plot



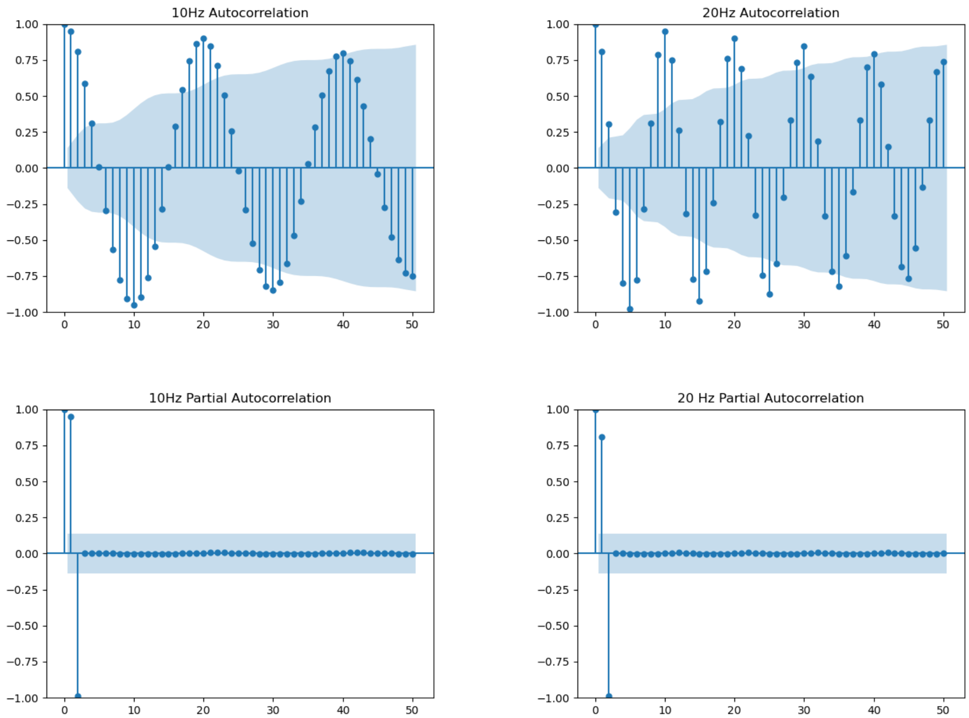
2.1.2 Power spectrum



2.1.3 Spectrogram

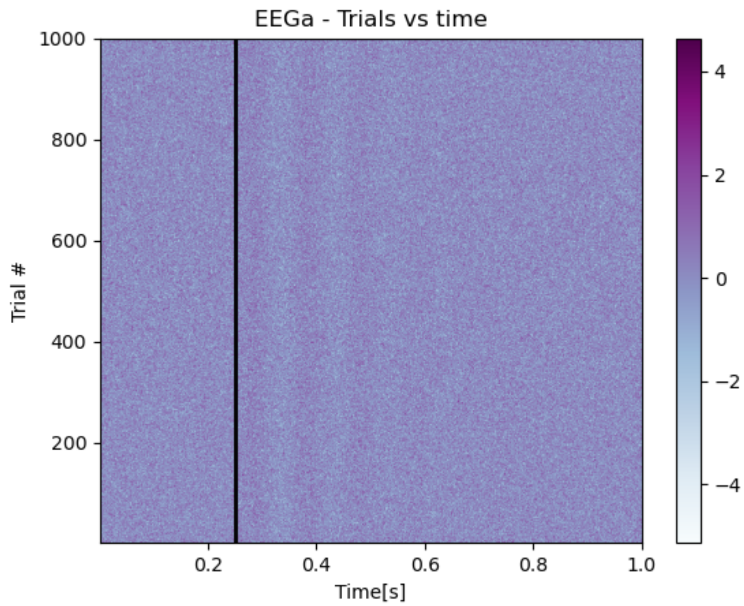


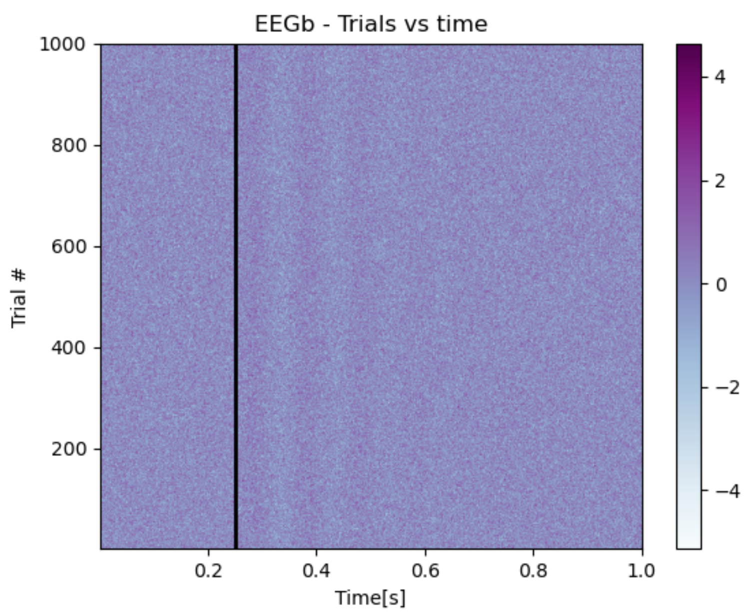
2.1.4 Compare ACF and PACF graphs

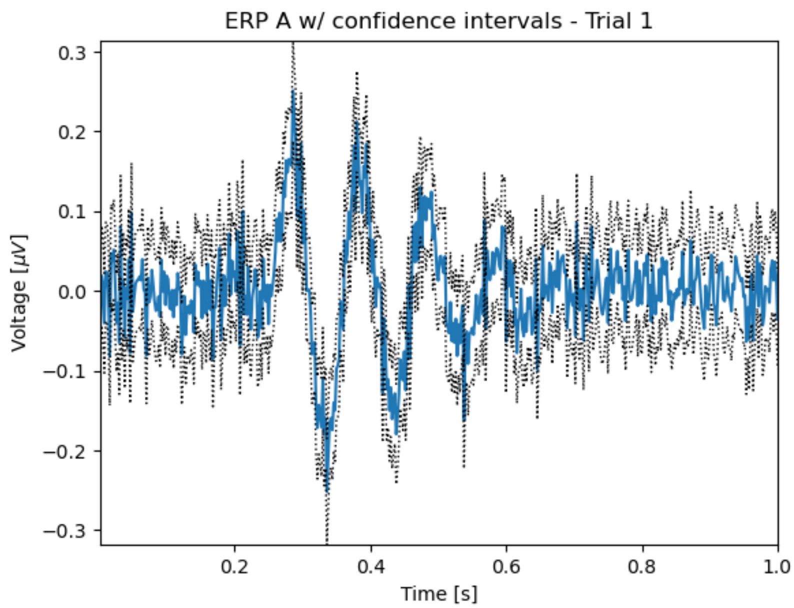


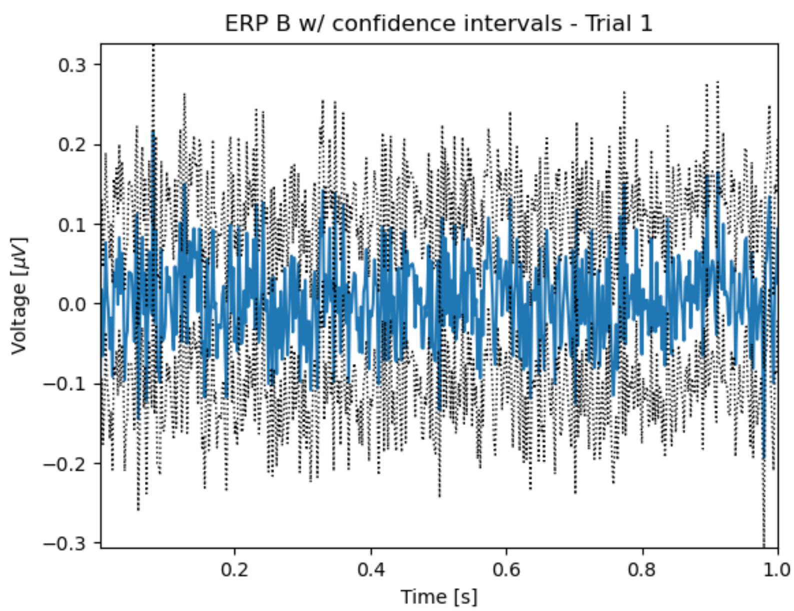
2.2 Statistical Features and discovery of event-related potential

2.2.1 Visualizing the response in conditions A and B

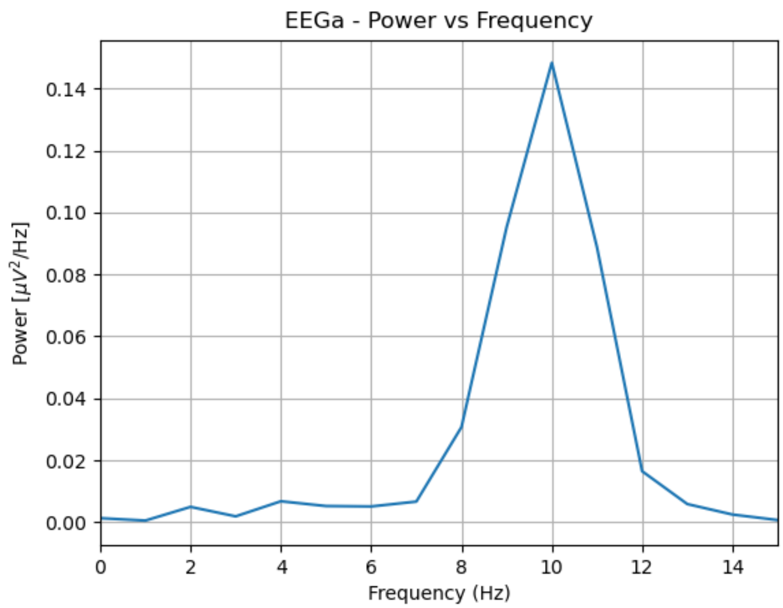





ERP A w/ confidence intervals - Trial 1
Voltage [μV]
Time [s]

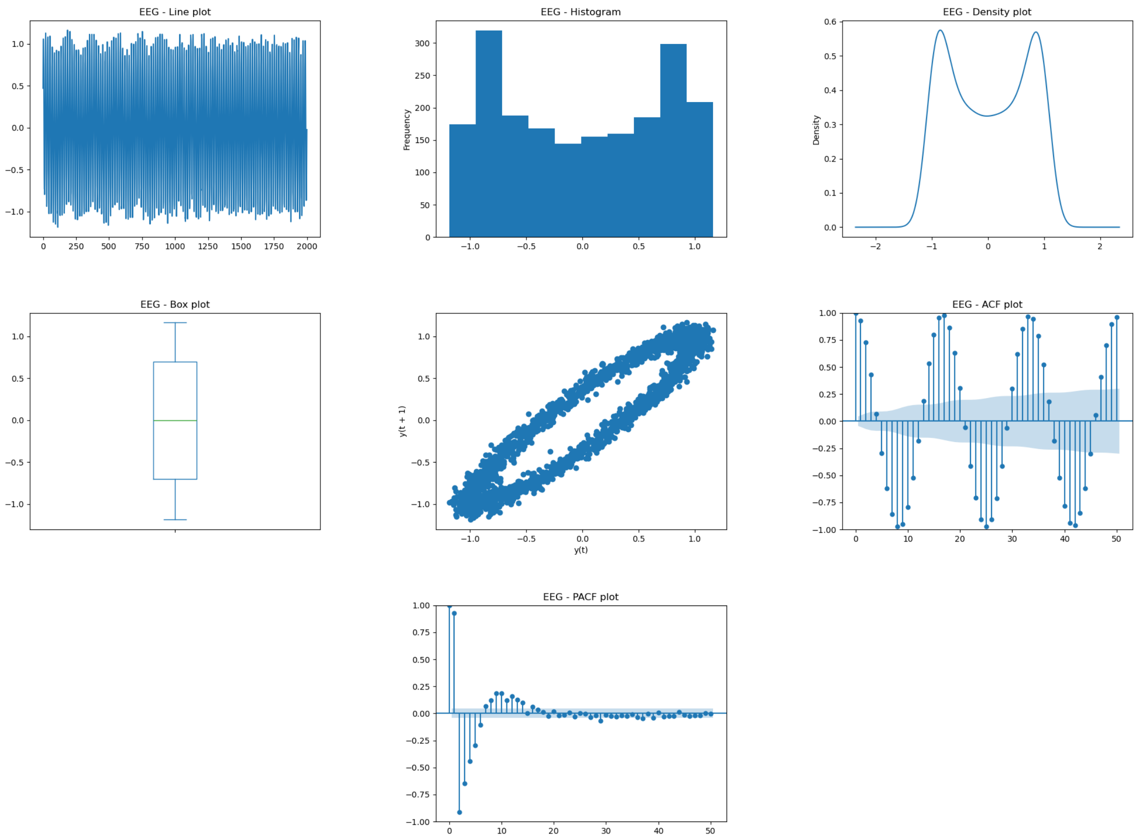

ERP B w/ confidence intervals - Trial 1
Voltage [μV]
Time [s]

2.2.2 Find brain activity frequency in data of condition A



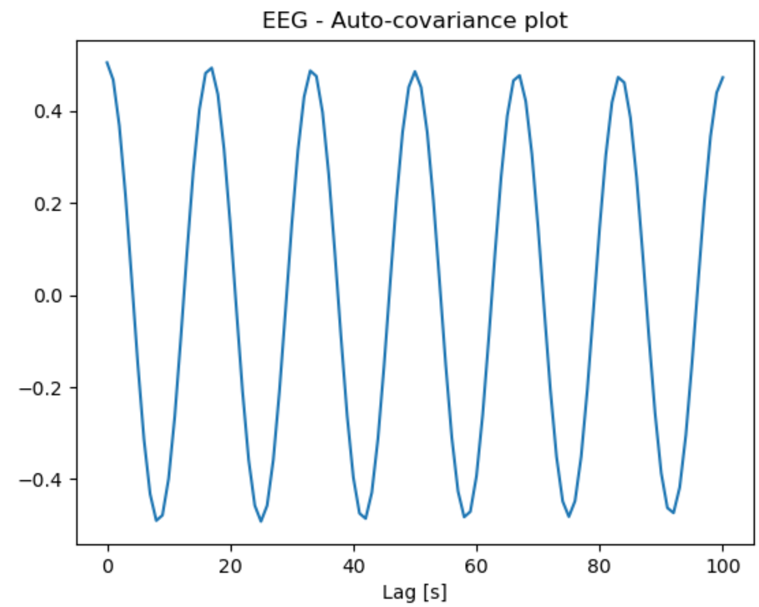
2.3 Features of observed rythms in EEG

2.3.1 Plots of EEG data



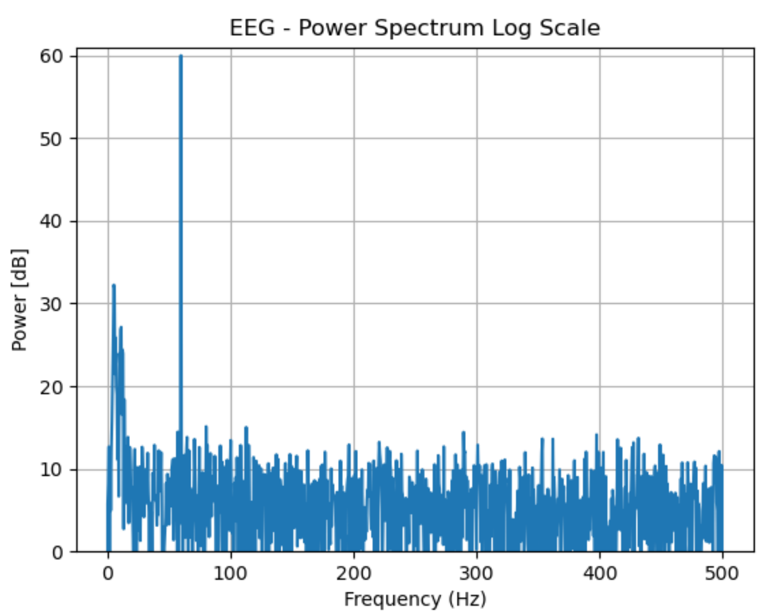
2.3.2 Statistical characterisitics of EEG data

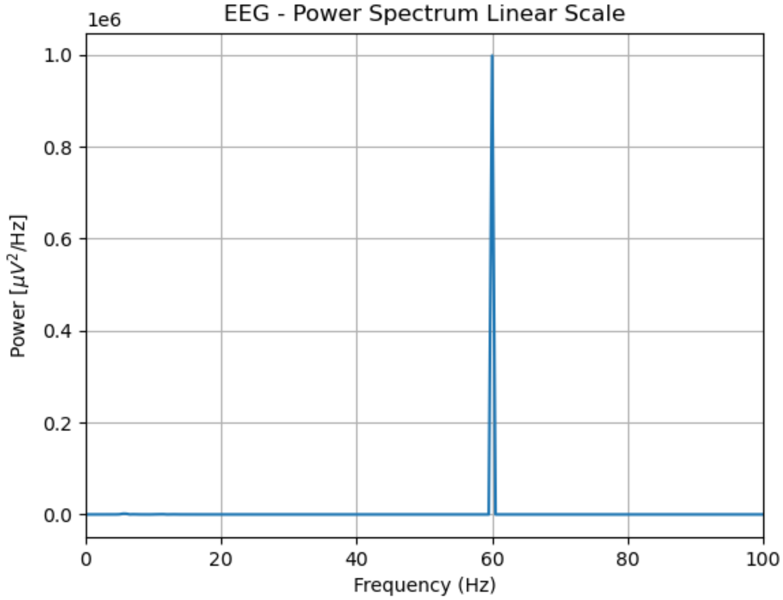
2.3.3 Auto-covariance of EEG data



Mean	2.7311486e-17
Variance	0.5049697256
Standard-deviation	0.71061221889

2.3.4 Power-spectrum of EEG data





2.3.5 Questions

What features do you typically consider useful for analyzing and modeling time-series data?

Features that I would consider important for analzying and modeling time series data are statistical features, temporal features and spectral features (Source: Lecture 3)

What features are specific for time-series, and what are general for both time-series and non-time-series data?
Features common to both time and non-time series are:

- Trend analysis
- Mean, Std, Variance
- Autocorrelation
- Stationarity

Features specific to time series are:

- Seasonality
- Residual analysis

How are auto-covariance and auto-correlation are defined for a time series? Give mathematical formulas for the definitions.

Biased auto-covariance:

$$\gamma_k = \frac{1}{N}\sum_{n=1}^{N-k}(X_n - \mu)(X_{n+k} - \mu)$$

Unbiased auto-covariance

$$\gamma_k = \frac{1}{N-k}\sum_{n=1}^{N-k}(X_n - \mu)(X_{n+k} - \mu)$$

Auto-correlation:

$$\rho_k = \frac{\gamma_k}{\gamma_0}$$

Assume a short time-series 1, 2, 3, 4, 5, 6, 7, 8, 7, 6, 5, 4, 3, 2, 1. (1) Calculate the auto-covariance and auto-correlations for all valid lags. Do the calculations manually. (2) Write a Python program to validate your calculations. (3) Draw the ACFgraph for the time series.

1. Compute Manually:

Step 1: Compute the mean and variance of the time series:

$$N = 15$$

$$\mu = \frac{1}{N}\sum_{n=1}^{N} X_n = \frac{64}{15} = 4.26$$

Step 2: Calculate the auto-covariance for each lag k:

$$\gamma_k = \frac{1}{N}\sum_{n=1}^{N-k}(X_n - \mu)(X_{n+k} - \mu)$$

At k = 0,

$$\gamma_0 = 4.72$$

At k = 1,

$$\gamma_0 = 3.55$$

Step 3: Compute the auto-correlation for each lag:

$$\rho_k = \frac{\gamma_k}{\gamma_0}$$

At k = 0,

$$\rho_1 = \frac{4.72}{4.72} = 1$$

At k = 1,

$$\rho_1 = \frac{3.55}{4.72} = 0.75$$

2. Python generated:

a. Auto-covariance: [4.72888889 3.55081481 2.07496296 0.50133333 -0.97007407 -2.13925926 -2.80622222 -2.77096296 -1.83348148 -0.93155556 -0.13185185 0.49896296 0.89422222 0.98725926 0.71140741]

b. Auto-correlation: [1. 0.75087719 0.43878446 0.10601504 -0.20513784 -0.45238095 -0.59342105 -0.58596491 -0.3877193 -0.19699248 -0.02788221 0.10551378]

Comparison of manual and with python calculated auto-covariance:

	Manual	Python
k_0	4.72	4.73
k_1	3.55	3.55

Comparison of manual and with python calculated auto-correlation:

	Manual	Python
k_0	1	1
k_1	0.75	0.75

3. ACF Plot