Techniques to ‘stationarize’ Time Series

Aria Alinejad, Cameron Louis Penne, Giacomo Melloni, Johannes Djupesland, Niclas Flehmig, Simen Dymbe, Thale Eliassen Fink

Contents

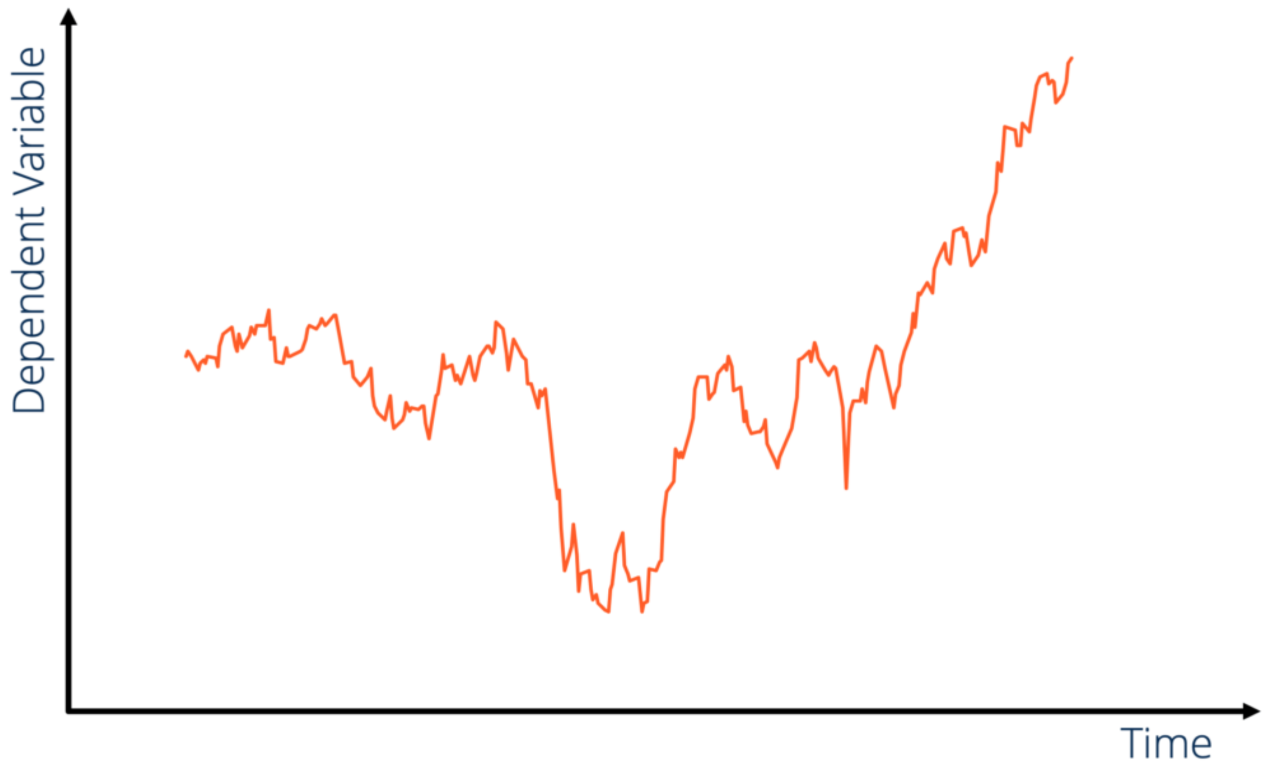
- Motivation
- What is stationarity
- Tests
- Methods for stationarizing
- Code example
- Conclusion

Making Time Series Stationary

MOTIVATION

Time Series

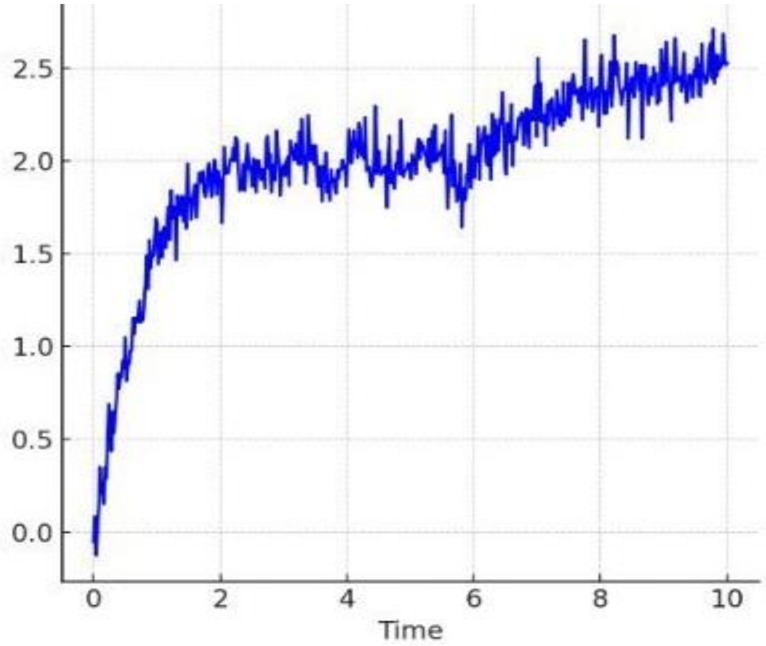
- A set of regularly or irregularly taken time-ordered numerical observations of some phenomena

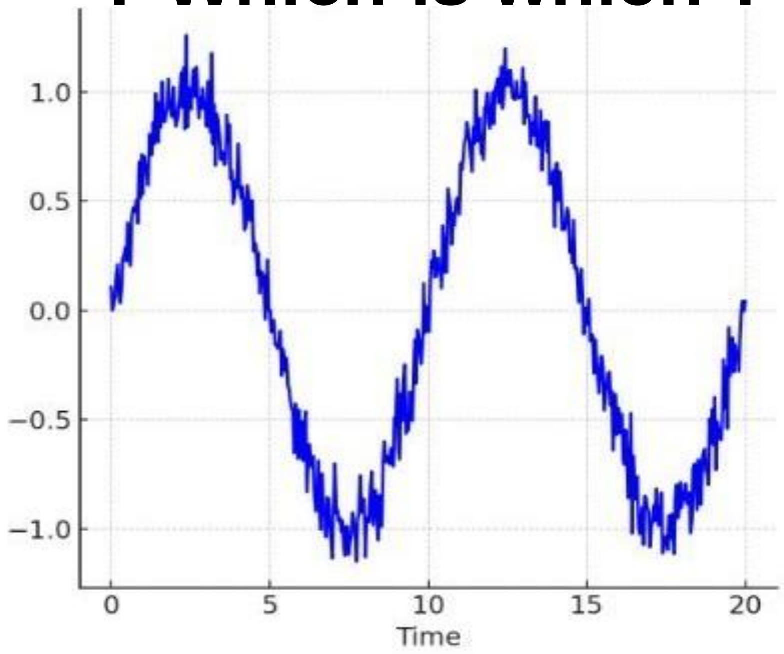


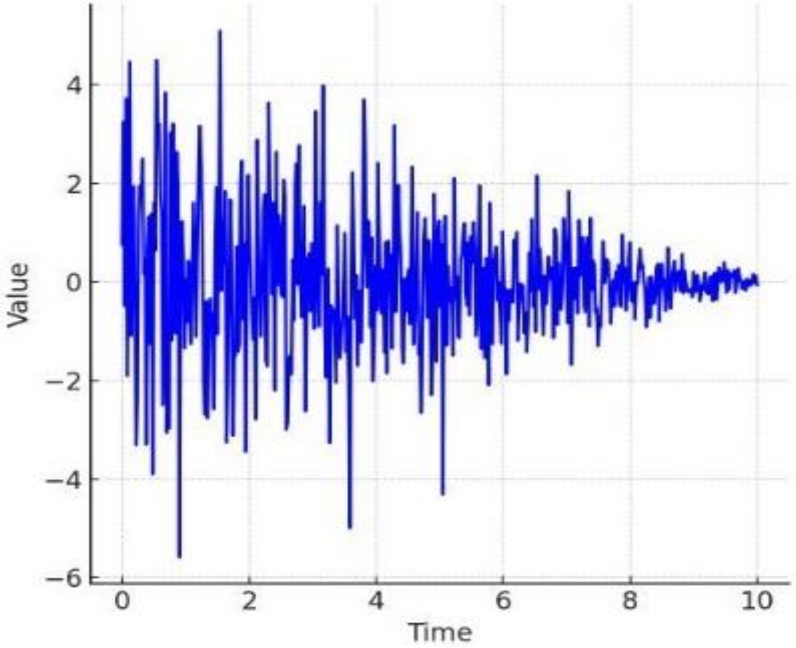
What is Stationarity?

- The statistical properties of a process generating a time series do not change over time.
  - Mean, Variance, Seasonality/trends

? Which is which ?



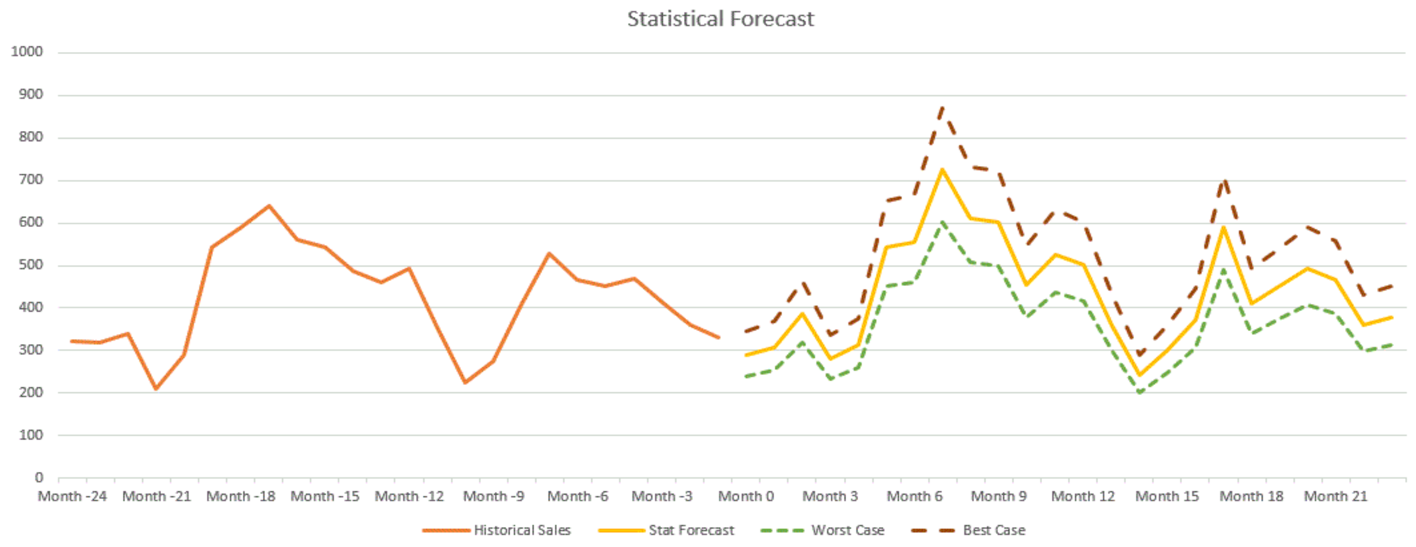




Towards Data Science: Stationarity in time series analysis (For more info, e.g. N'th order, Weak and strong stationarity

Why do we need Stationarity?

- **Statistical Forecasting**
  - Process of predicting future event. Assumes stationary data
- **Modeling Assumptions**
  - Many time series models require stationary data (e.g. ARIMA, SARIMA)



How to know if your series is stationary?

TEST FOR STATIONARITY

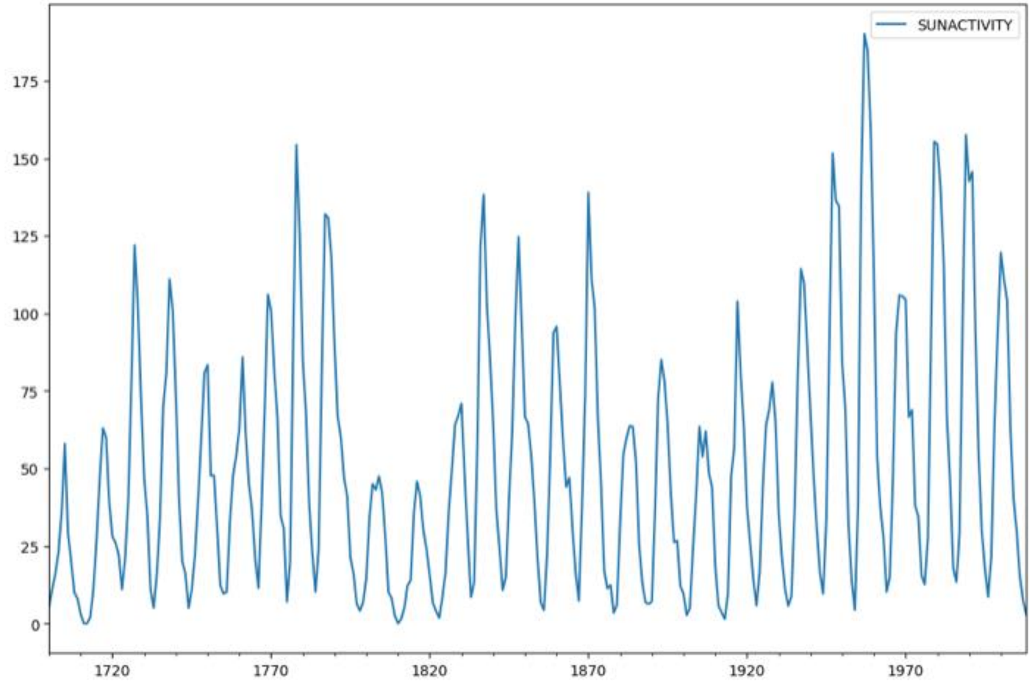
Common tests for stationarity

- **Statistical tests** to check for (trend-)stationarity
- **Hypothesis testing** with null and alternate hypothesis
- Perform **inference** based on test statistic and p-values to decide on hypothesis test
- p-value must be below significance level (e.g. 0.05) to reject null hypothesis
- Existence of **unit root** which indicates non-stationarity
- **Time lag** refers to time interval between observations in a dataset

Augmented Dickey-Fuller (ADF)

- Check for stationarity in data
- **Null hypothesis:** *Series has a unit root.* -> non-stationary
- **Alternate hypothesis:** *Series has no unit root.* -> stationary
- **Intuition:**
  - **If unit root:** the lagged observations do not provide relevant information to me
  - **If no unit root:** the lagged observations provide relevant information to me

Example



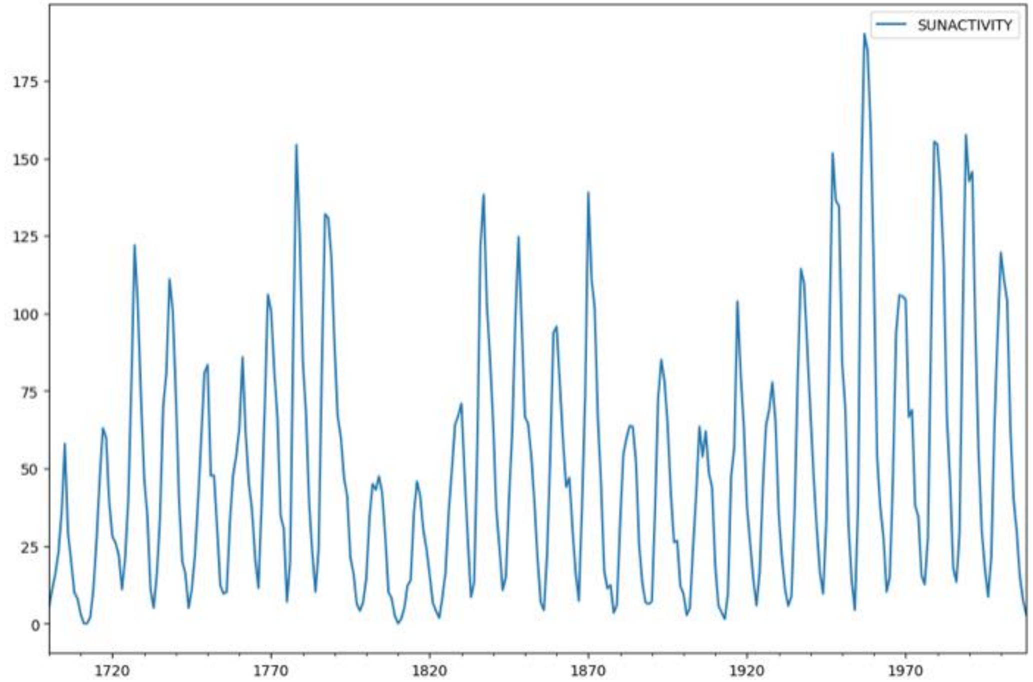

```
Results of Dickey-Fuller Test:
Test Statistic                  -2.837781
p-value                          0.053076
#Lags Used                       8.000000
Number of Observations Used    300.000000
Critical Value (1%)             -3.452337
Critical Value (5%)             -2.871223
Critical Value (10%)            -2.571929
dtype: float64
```

- p-value is larger than significance level (0.05)
- Used 8 lags
- Critical value represents the confidence interval
- Use test statistic to compare to critical values

Kwiatkowski-Phillips-Schmidt-Shin (KPSS)

- Check for stationarity in presence of deterministic trend
  - ***'deterministic trend'*** – slope of trend does not change permanently
- **Null hypothesis:** *Process is trend stationary.*
- **Alternate hypothesis:** *Series has a unit root.* -> non-stationary

Example




```
Results of KPSS Test:
Test Statistic            0.669866
p-value                   0.016285
Lags Used                 7.000000
Critical Value (10%)      0.347000
Critical Value (5%)       0.463000
Critical Value (2.5%)     0.574000
Critical Value (1%)       0.739000
dtype: float64
```

- p-value is smaller than significance level
- Reject the null hypothesis

Application

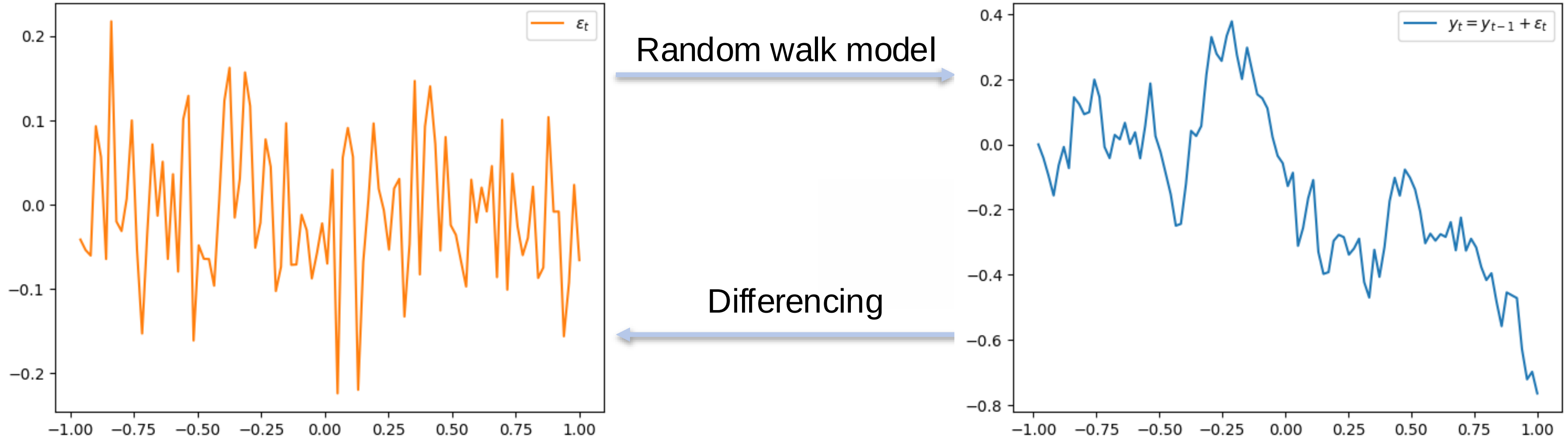
- ADF and KPSS should not be used interchangeable
- Rejecting null hypothesis has almost opposite meaning in ADF and KPSS
- Good practice to always use both to ensure stationarity
  - ***Case 1:*** both tests conclude non-stationarity -> non-stationarity
  - ***Case 2:*** both tests conclude stationarity -> stationarity
  - ***Case 3:*** KPSS stationarity, ADF non-stationarity -> trend stationarity, de-trending is to be used
  - ***Case 4:*** KPSS non-stationarity, ADF stationarity -> difference stationarity, differencing is to be used

Stationarizing Time Series

METHODS

Differencing

- Assume a random walk model: $y_t = y_{t-1} + \varepsilon_t$
  - Where $\varepsilon_t \sim some\ distrubtion(with\ constant\ parameters)$



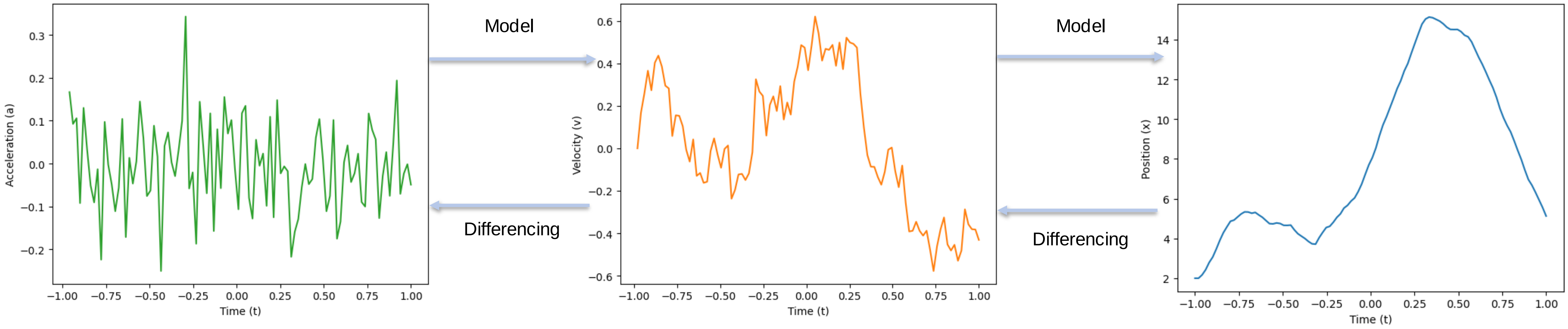
- Taking the difference between samples gives us a stationary time-series:

$$\varepsilon_t = y_t - y_{t-1}$$

Second degree differencing:

- Ex: modelling position with random acceleration:

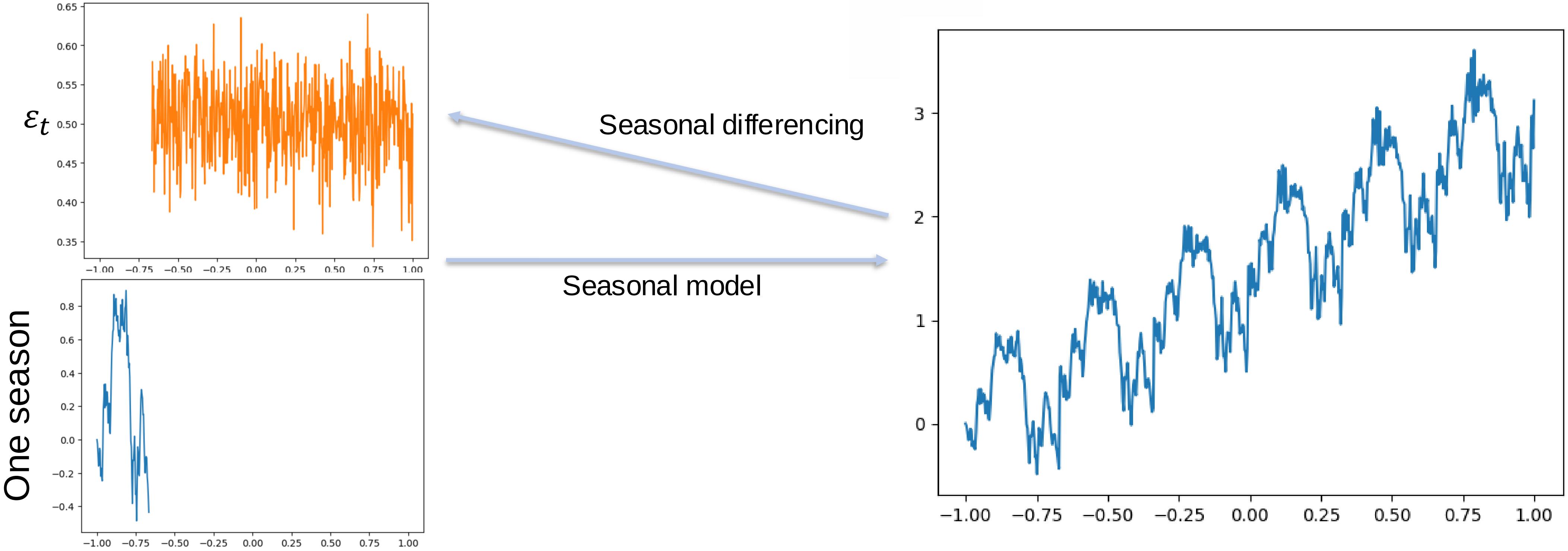
$$x_t = x_{t-1} + v_t$$
$$v_t = v_{t-1} + a_t$$
$$a_t \sim some\ distrubtion(with\ constant\ params)$$



Differencing twice!

Seasonal differencing

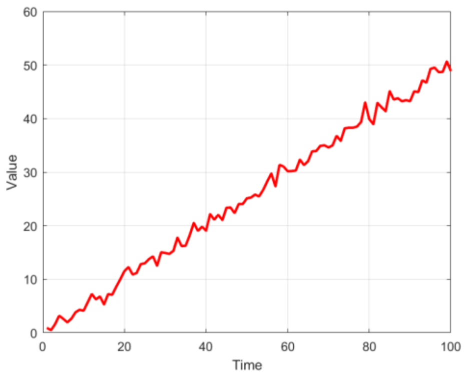
- Assume seasonal model: $y_t = y_{t-m} + \varepsilon_t$
  - Where m is the period in samples of the seasonal effect
  - $\varepsilon_t \sim some\ distrubtion(with\ constant\ parameters)$
- Seasonal differencing: $\varepsilon_t = y_t - y_{t-m}$



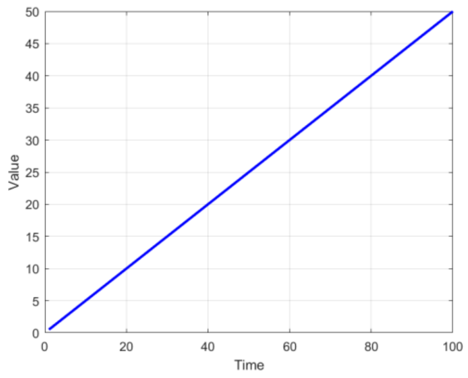
De-trending

- Identify and remove long-time trends in time series
- Common trend identification-methods:
  - Regression
  - Moving average smoothing

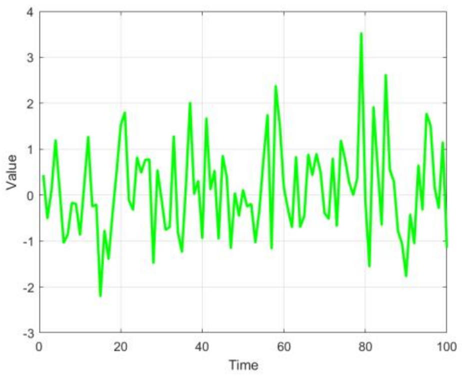
Original data



Long-time trend



Original data – Long time trend

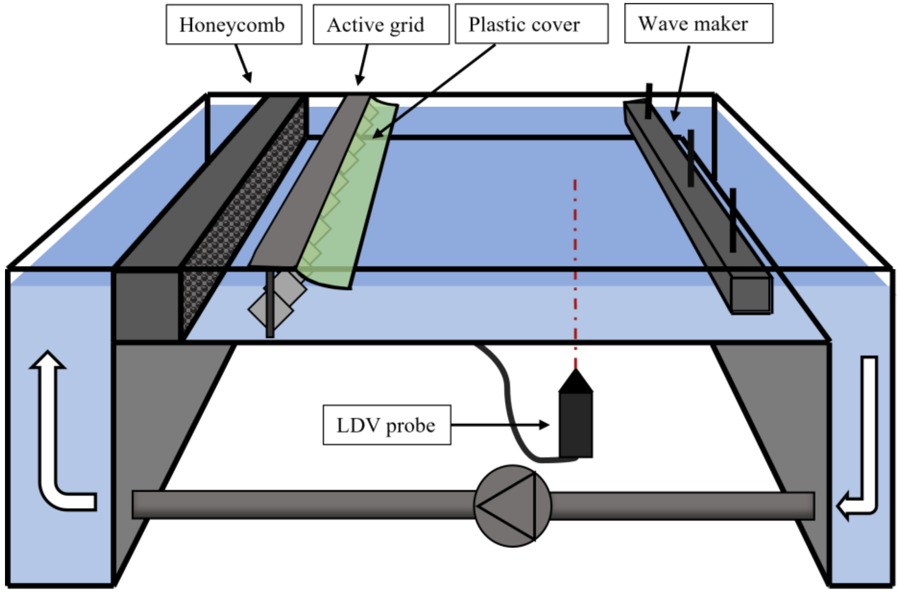


Seasonal adjustments

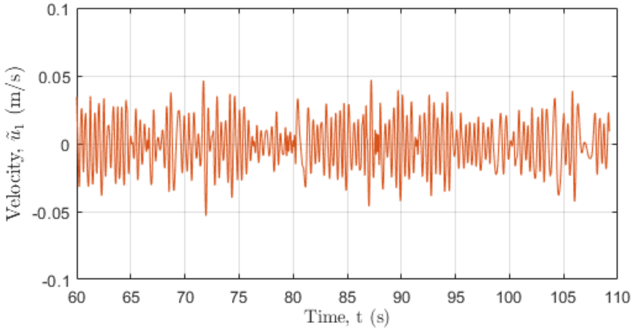
- Identify and remove repeating patterns in a time series
- Common methods:
  - Decomposition
    - EMD
    - Wavelet transform
    - FFT
  - Moving average smoothing

Decomposition example

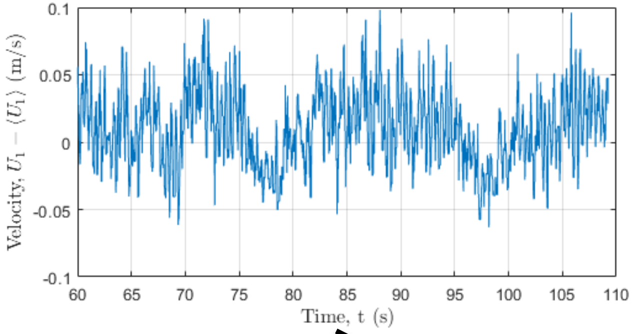
- Surface wave-turbulence interaction



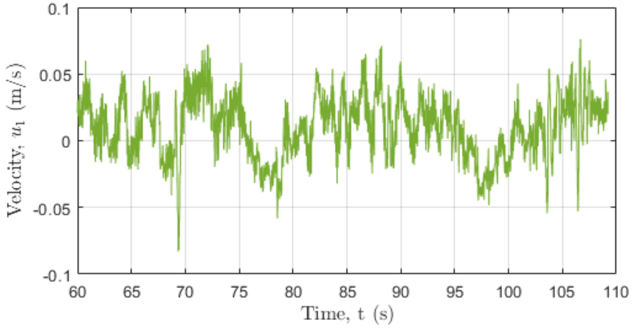
Wave component



Raw data

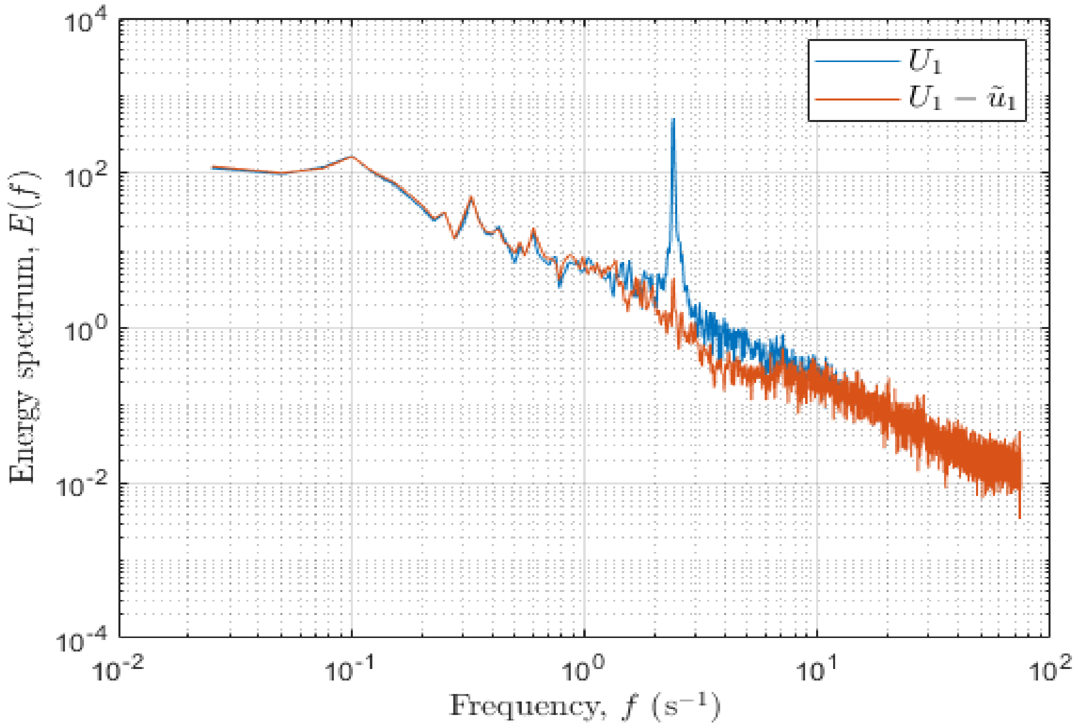


Turbulence component

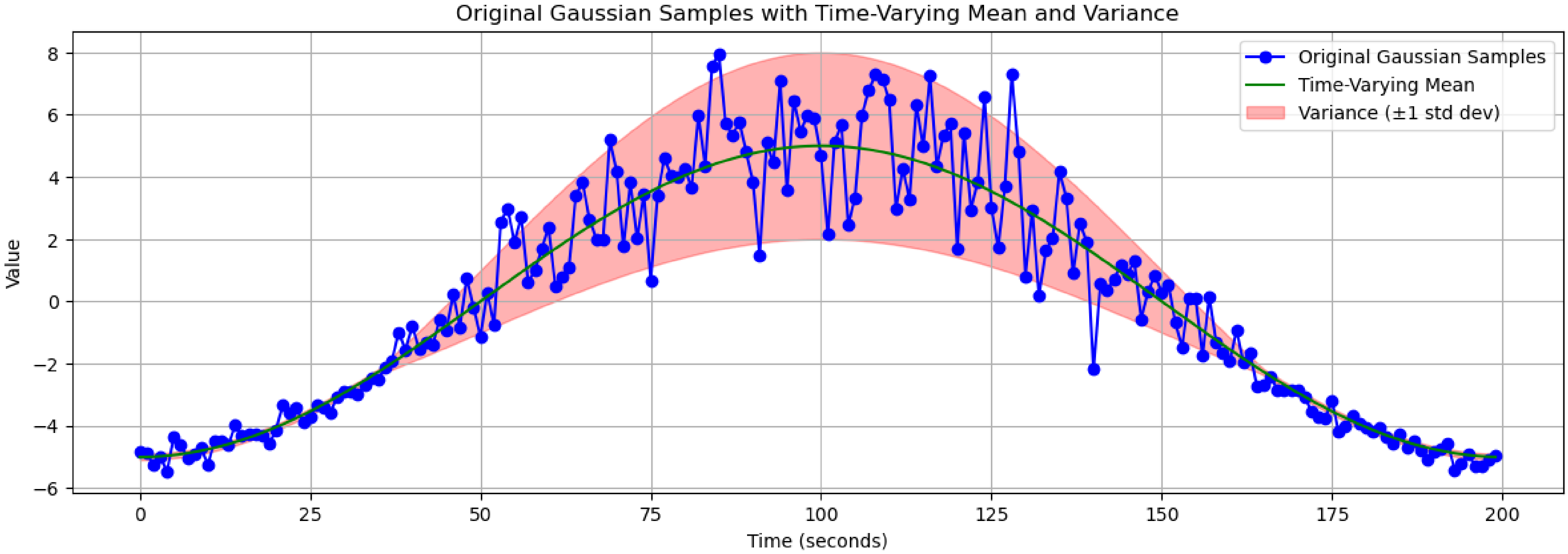


Decomposition example

- Not a perfect decomposition
  - Did not completely remove wave motion
  - Removed some of the turbulence motion



Code Example - Dataset



Stationarity tests

ADF test:

- p-value: 0.775
- ADF statistic: -0.938
- Critical values:

1%	5%	10%
-3.5	-2.9	-2.6

H0: Non-stationary

- The p-value is above 0.05
  → fail to reject H0
- ADF statistic > critical values of all significance levels
  → fail to reject H0

KPSS test:

- p-value: 0.036
- KPSS statistic: 0.523
- Critical values:

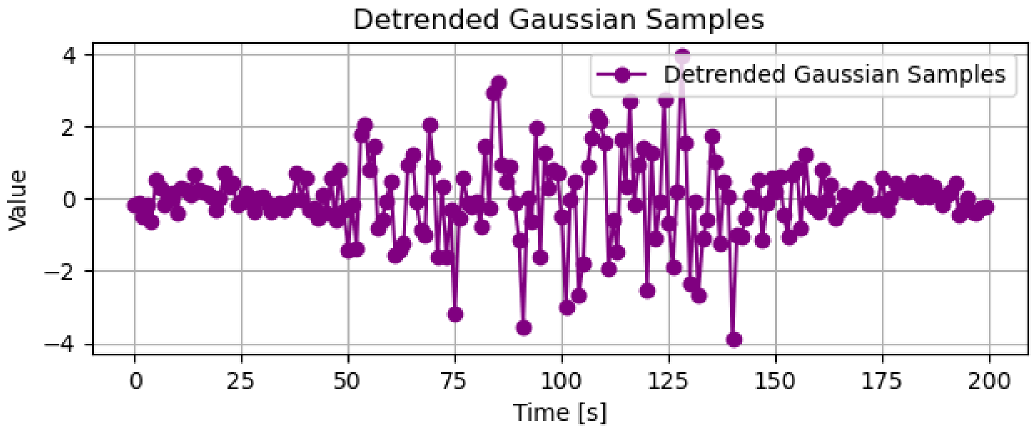
1%	2.5%	5%	10%
0.74	0.574	0.46	0.35

H0: stationary

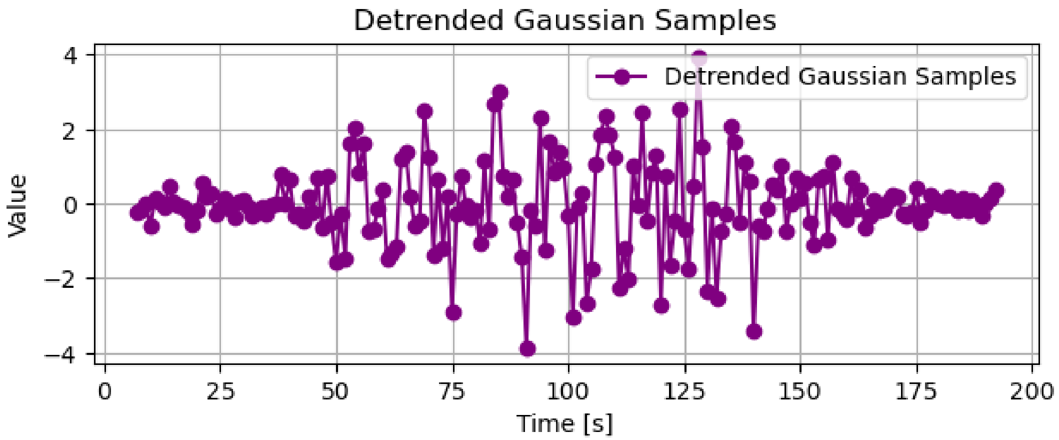
- The p-value is below 0.05
  → reject H0
- 5% < KPSS statistic < 2.5%
  → reject H0 at 5% significance

Detrending methods

Detrending with Moving average

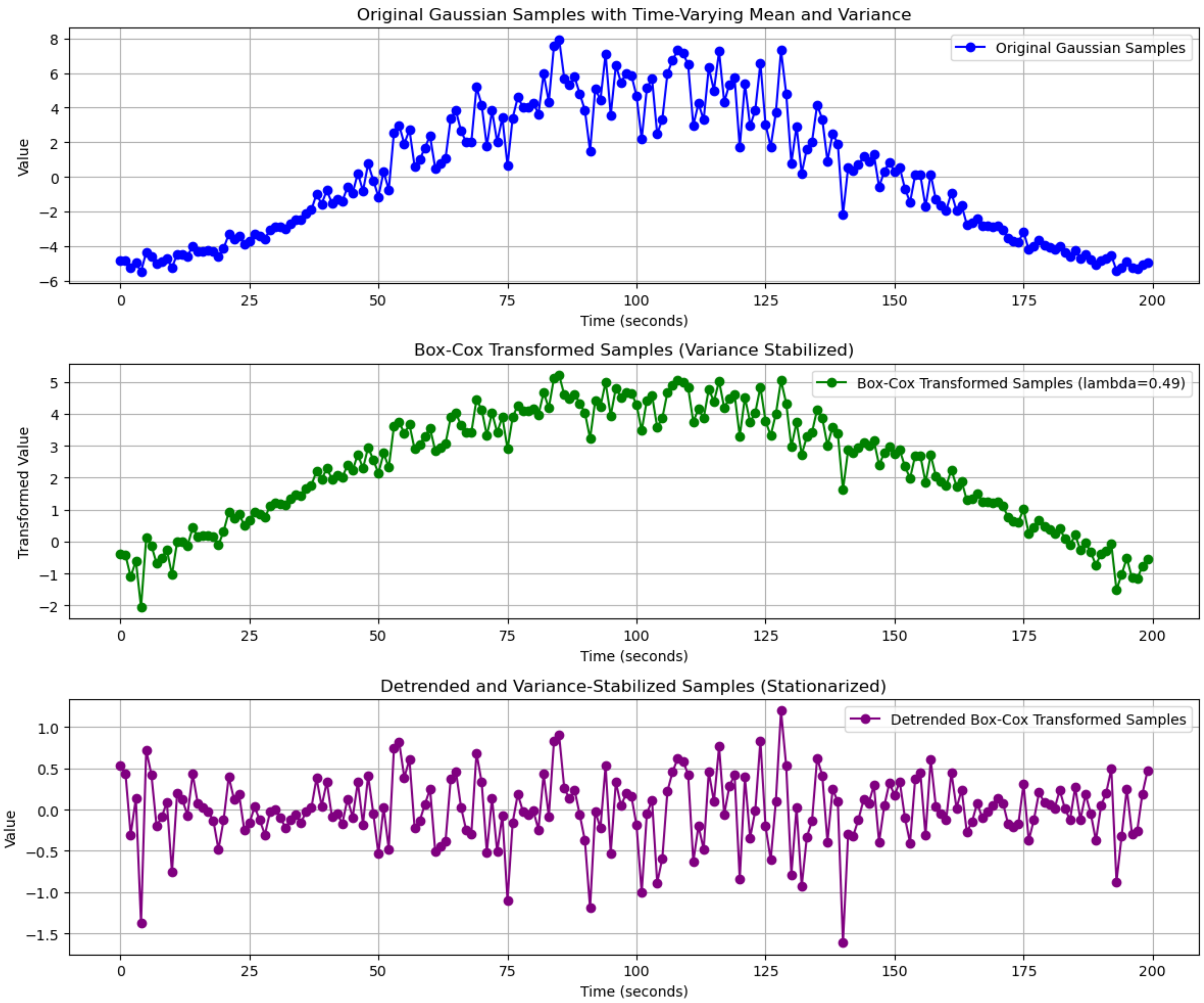


Detrending with Polynomial regression (order=4)



Mean is adjusted for, but both still have varying std

Detrending with Box Cox



1. Box cox adjusts variance
2. Detrend adjusts mean

Stationarity tests AGAIN!

ADF test:

- p-value: $3.84 * 10^{-26}$
- ADF statistic: -14.002
- Critical values:

1%	5%	10%
-3.5	-2.9	-2.6

H0: Non-stationary

- The p-value is below 0.05
  → reject H0
- ADF statistic < critical values of all significance levels
  → reject H0

KPSS test:

- p-value: 0.1
- KPSS statistic: 0.0339
- Critical values:

1%	5%	10%
0.74	0.46	0.35

H0: stationary

- The p-value is above 0.05
  → keep H0
- KPSS statistic < critical values of all significance levels
  → keep H0

CONCLUSION

Conclusion

Why:

- Want to model/predict $\longrightarrow$ need stationarity

How to check:

- Visually
- ADF
- KPSS

How to "fix":

- Differencing
- De-trending
- Seasonal adjustment
- Box Cox + De-trending
- Other..

Last slide ☺