HSLU Hochschule Luzern

Time Series Analysis

Foundations I

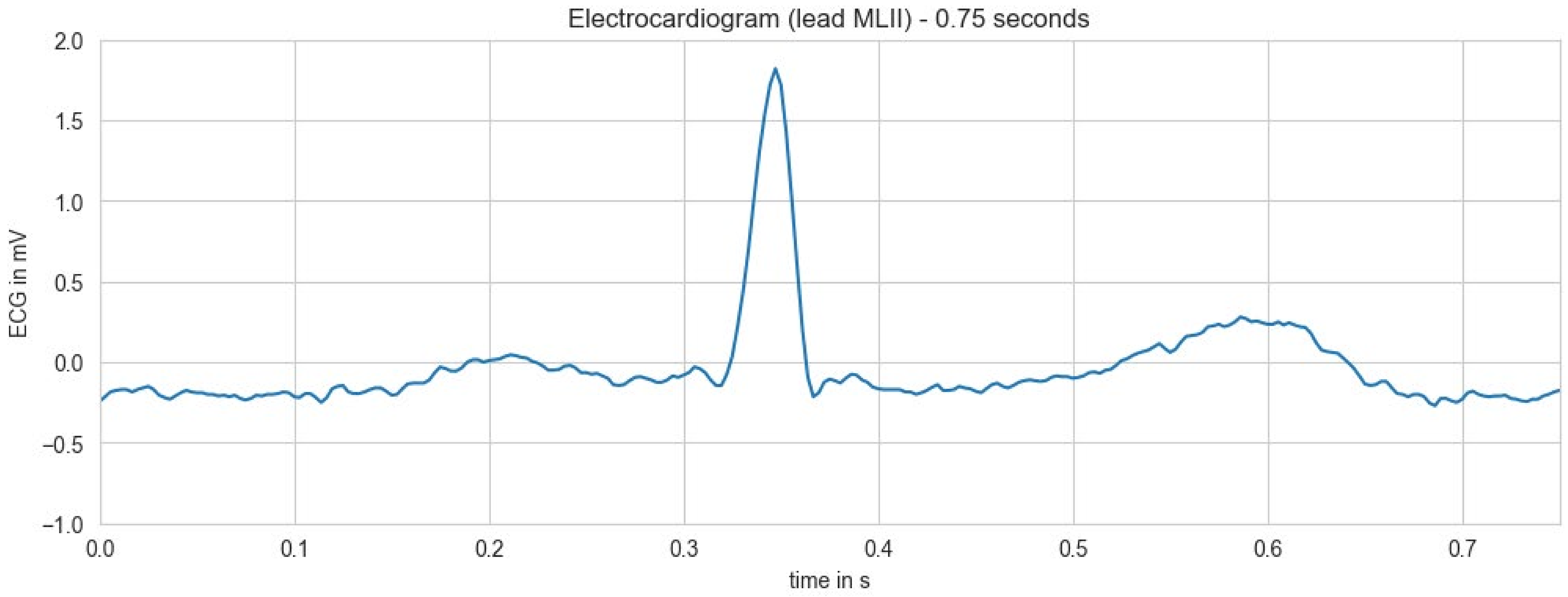
Dr. Ludovic Amruthalingam
ludovic.amruthalingam@hslu.ch

Informatik

Outline

- Statistical model for time series
- Auto-correlation
- White noise
- Signal + noise model
- Time series decomposition
- Random walks
- Moving average smoothing
- Period adjusted average
- STL decomposition

Time series examples – ECG signal



Moody et al. The impact of the MIT-BIH Arrhythmia Database. IEEE Eng in Med and Biol 20(3):45-50 (May-June 2001). (PMID: 11446209); DOI:10.13026/C2F305

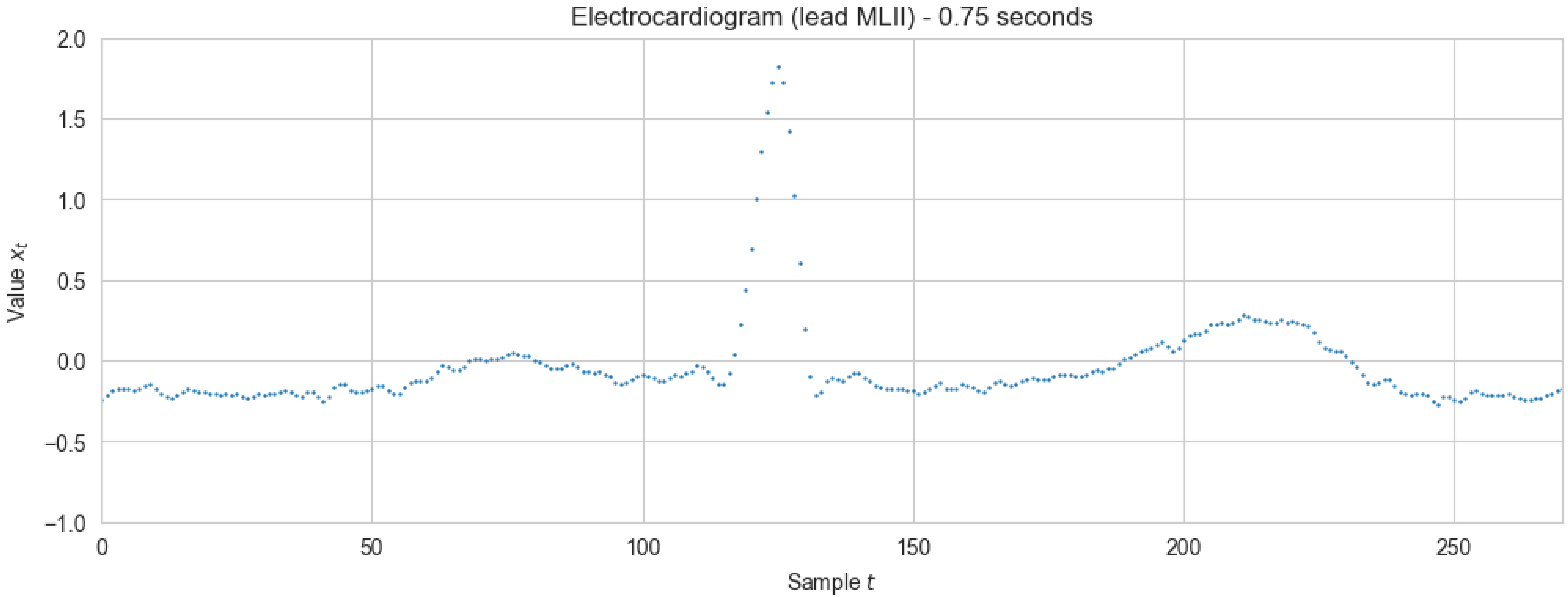
Statistical model for time series

A time series (TS) is a collection of **data points** observed **sequentially in time**.

Model TS as **stochastic processes**, i.e., collections of **random variables** (RV), $\{X_1, X_2, \dots X_n\}$, indexed according to their observation order.

A model specifies the **joint distribution** of the sequence of RVs $P[X_1 \leq x_1, \dots X_n \leq x_n]$ in the continuous case, $P[X_1 = x_1, \dots X_n = x_n]$ in the discrete case, where $\{x_1, x_2, \dots x_n\}$ is a **realization** of the stochastic process.

A realization of a stochastic process



ecg: $[-0.245, \quad -0.215, \quad -0.185, \quad -0.175, \quad -0.17, \quad -0.17, \quad -0.185, \quad -0.17, \quad -0.16, \quad -0.15, \quad \ldots]$

ecg: $[x_1, \quad x_2, \quad x_3, \quad x_4, \quad x_5, \quad x_6, \quad x_7, \quad x_8, \quad x_9, \quad x_{10}, \quad x_{11}, \quad \ldots]$

Moody et al. The impact of the MIT-BIH Arrhythmia Database. IEEE Eng in Med and Biol 20(3):45-50 (May-June 2001). (PMID: 11446209); DOI:10.13026/C2F305

Recap – Probability concepts

Random variables (RV) are **real-valued functions** whose outcomes vary due to … randomness.

Property	**Expectation**	**Variance**	**Covariance**
Intuition	Long-term average, mean	Spread around mean	Joint variability, strength and direction of linear relationship
Formula	$E[X] = \sum_x x \cdot P(X = x) = \mu_X$	$Var(X) = E\left[(X - \mu_X)^2\right] = \sigma_X^2$	$cov(X, Y) = E[(X - \mu_X)(Y - \mu_Y)] = \gamma_{X,Y}$
Linearity	$E[aX + b] = aE[X] + b$	$Var(aX + b) = a^2 Var(X)$	$cov(aX + b, cY + d) = ac \cdot cov(X, Y)$
Additivity	$E[X + Y] = E[X] + E[Y]$	$Var(X + Y) = Var(X) + Var(Y) + 2cov(X, Y)$	$cov(X + Y, Z) = cov(X, Z) + cov(Y, Z)$

Note that $\gamma_{X,Y} = \gamma_{Y,X}$ (**symmetric**), that $\gamma_{X,Y} = E[XY] - \mu_X \mu_Y$ and that $\gamma_{X,X} = Var(X)$

We say that X and Y are **independent** $\Leftrightarrow P(X = x, Y = y) = P(X = x)P(Y = y) \Rightarrow \gamma_{X,Y} = 0$

The **correlation** is the normalized covariance: $\mathrm{cor(X, Y)} = \frac{\gamma_{X,Y}}{\sigma_X \sigma_Y} = \rho_{X,Y}$ with $-1 \leq \rho_{X,Y} \leq 1$

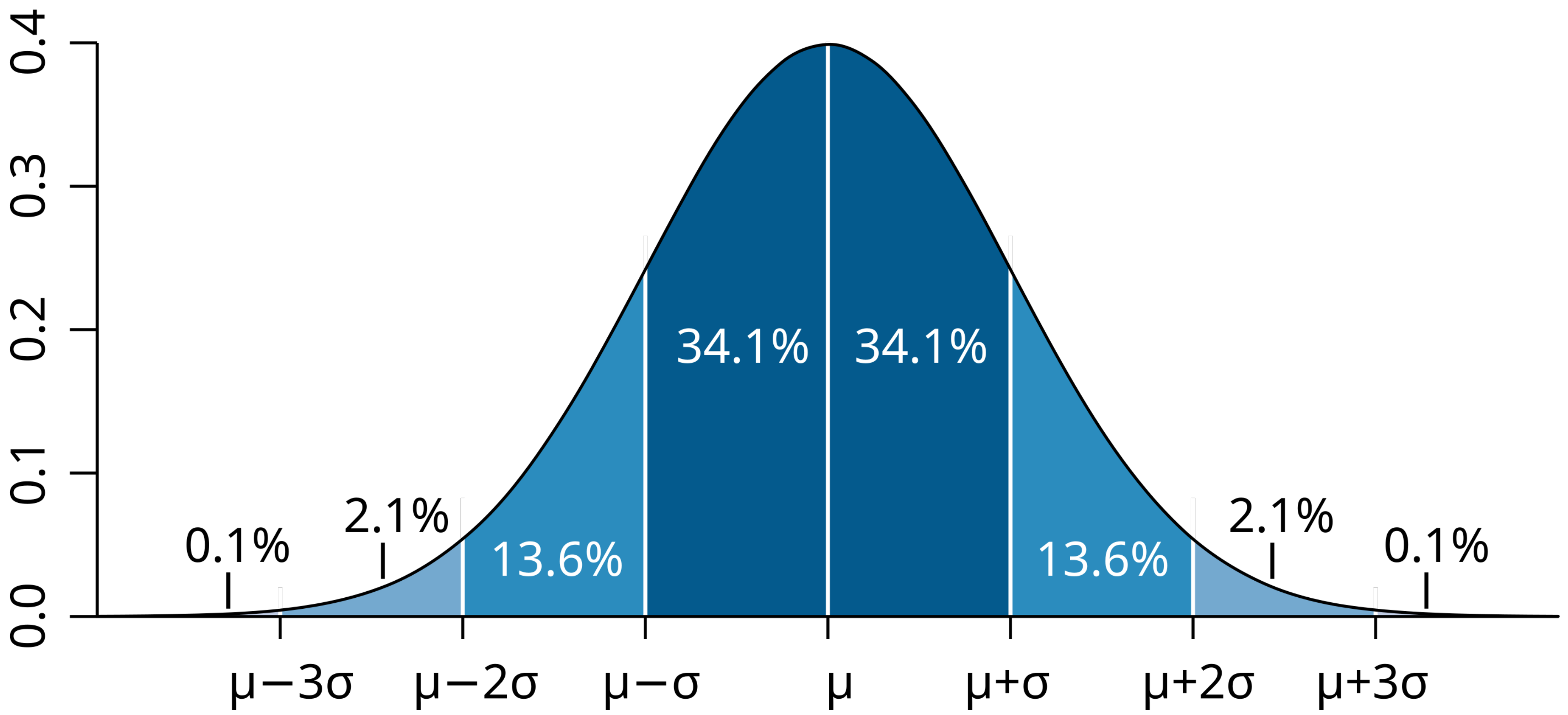
With TS, the **auto-covariance** is denoted $\gamma_{X_s,X_t} = \gamma(s,t)$ and the **auto-correlation** $\rho_{X_s,X_t} = \rho(s,t)$

covariance with it's shifted self

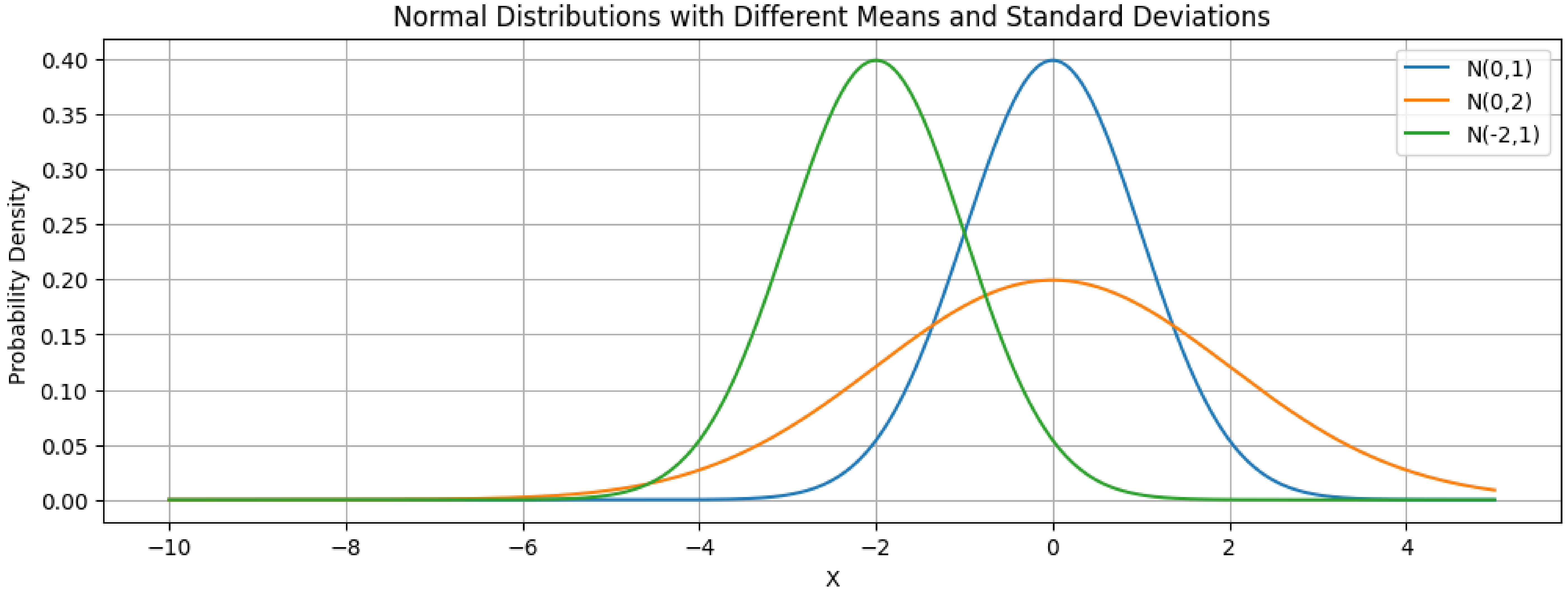
Recap – Probability concepts

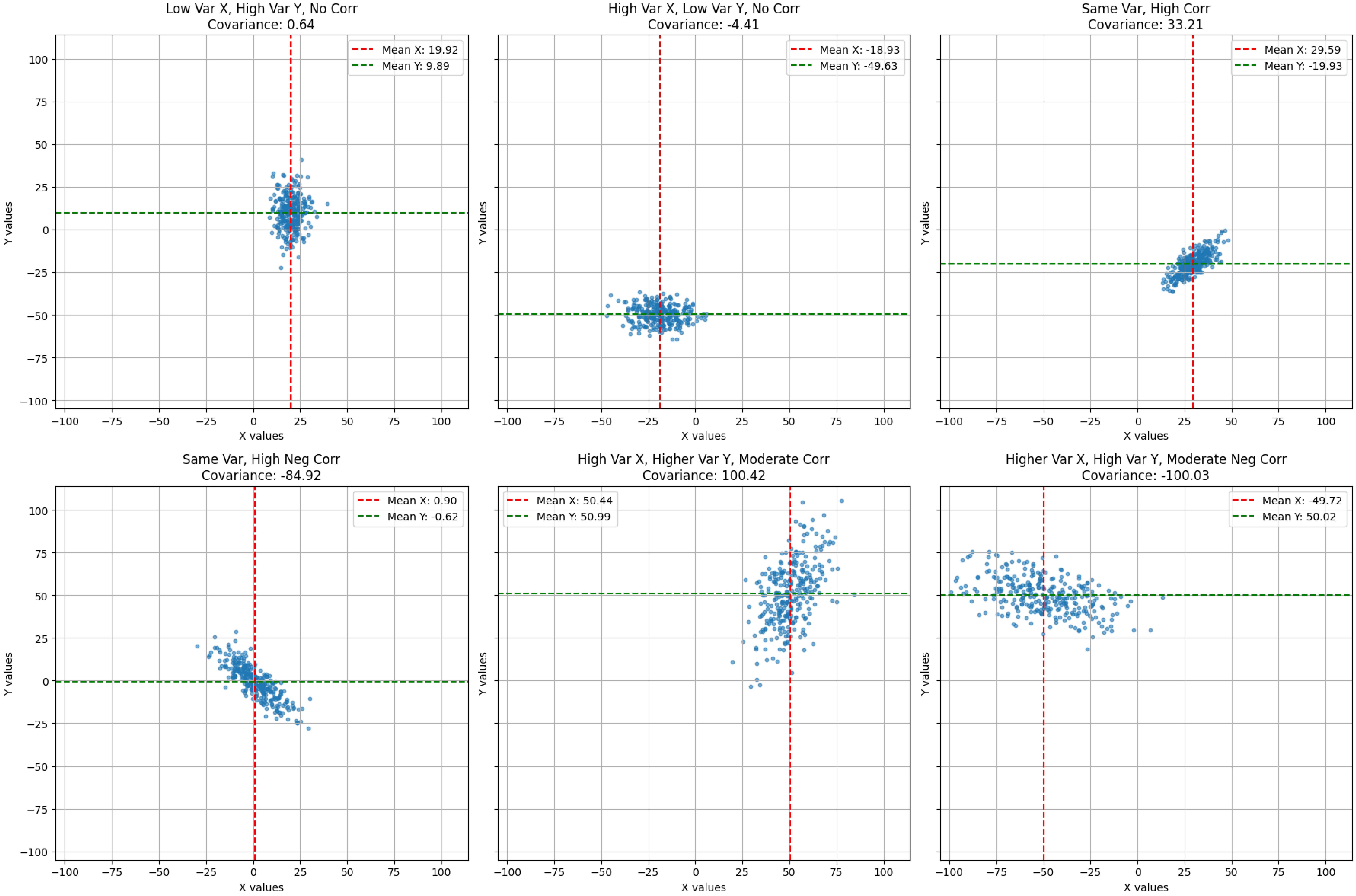
Probability distributions describe how probabilities are distributed over the values of the random variables.

Normal distribution $\mathcal{N}(\mu, \sigma^2)$: $f(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

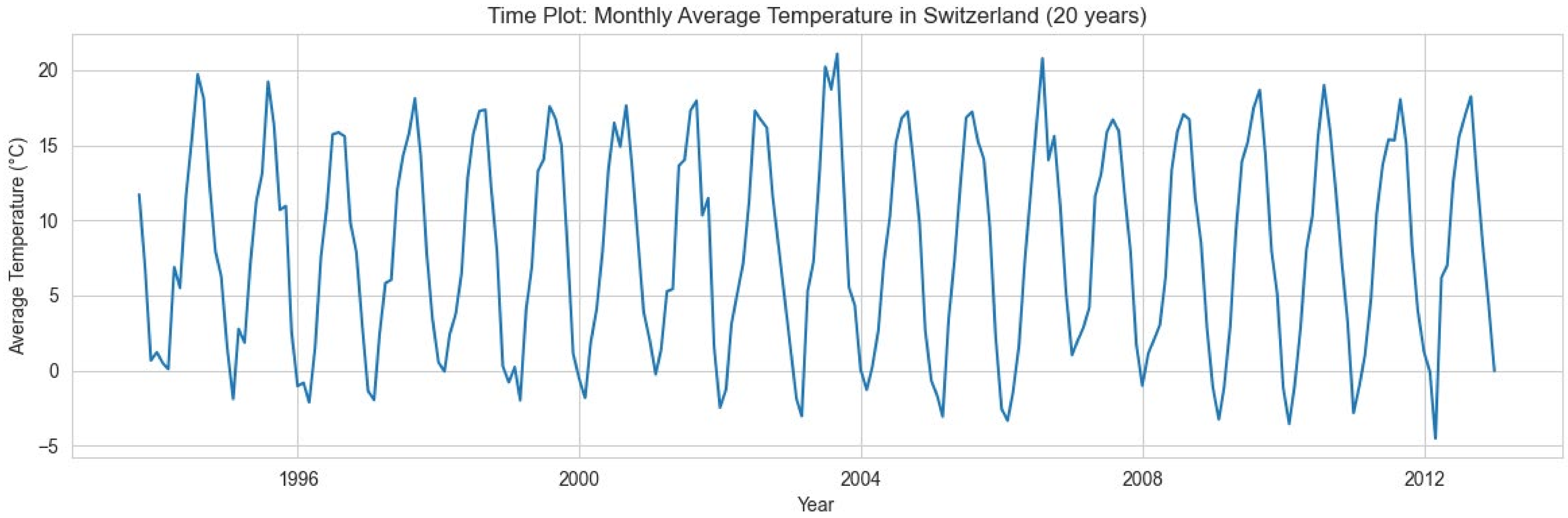


Ainali, Wikipedia



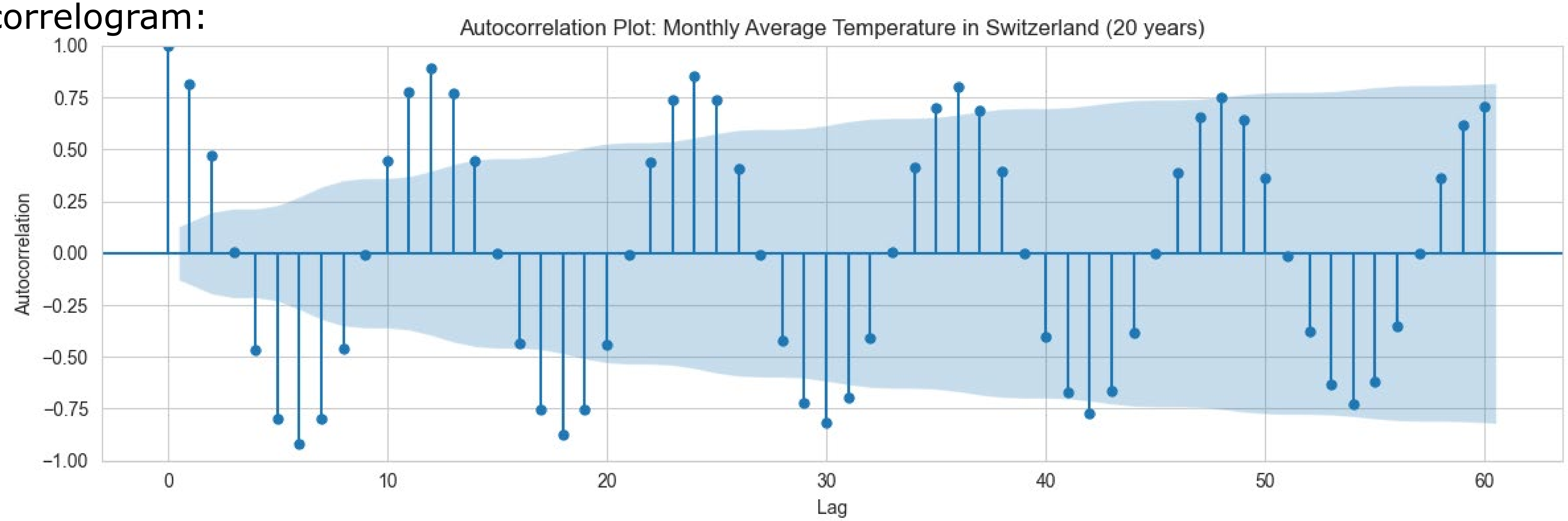

Low Var X, High Var Y, No Corr
Covariance: 0.64
Mean X: 19.92
Mean Y: 9.89
High Var X, Low Var Y, No Corr
Covariance: -4.41
Mean X: -18.93
Mean Y: -49.63
Same Var, High Corr
Covariance: 33.21
Mean X: 29.59
Mean Y: -19.93
Same Var, High Neg Corr
Covariance: -84.92
Mean X: 0.90
Mean Y: -0.62
High Var X, Higher Var Y, Moderate Corr
Covariance: 100.42
Mean X: 50.44
Mean Y: 50.99
Higher Var X, High Var Y, Moderate Neg Corr
Covariance: -100.03
Mean X: -49.72
Mean Y: 50.02
X values
Y values

The auto-correlation function (ACF) $\rho(s, t)$



ACF plot or correlogram:

$Lag = s - t$



Variance is constant over time

Confidence intervall

ignificantly different from 0

White noise (WN) process

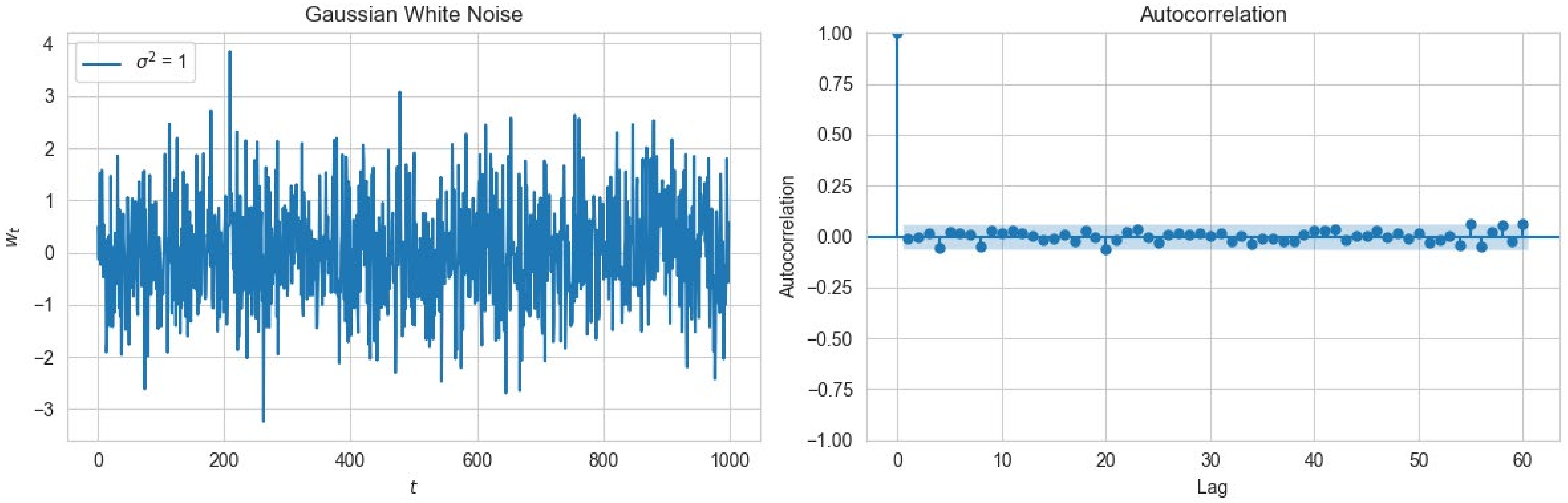
RVs $\{W_1, W_2, \ldots W_n\}$ are **uncorrelated** ($Cov(W_s, W_t) = 0, \forall s \neq t$), with **zero-mean** and **constant finite variance** σ^2:

$$W_t \sim WN(0, \sigma^2)$$

An additional assumption is that the RVs are **independent** and **identically distributed** $W_t \sim iid\ WN(0, \sigma^2)$:

$$P[W_1 = w_1, \ldots W_n = w_n] = P[W_1 = w_1] \ldots P[W_n = w_n]$$

Another useful assumption is that the RVs follow a **Gaussian** distribution (Gaussian WN) $W_t \sim \mathcal{N}(0, \sigma^2)$.



Signal + Noise model and time series decomposition

Signal + noise model: $X_t = \theta_t + E_t$ where θ_t is a modellable signal and E_t are errors (also called **innovations**).

Time series decomposition separates X_t into three components:

- Trend/trend-cycle U_t: increasing/decreasing pattern in the data.
- Seasonality S_t: repeating pattern with roughly fixed period.
- Remainder $\mathrm{R_t}$: everything remaining (**residuals**), E_t if θ_t is fully represented by U_t and S_t.

Additive decomposition $X_t = U_t + S_t + R_t$

- Magnitude of trend and seasonal fluctuations do not vary with the value of the TS.

Mutiplicative decomposition $X_t = U_t \times S_t \times R_t$

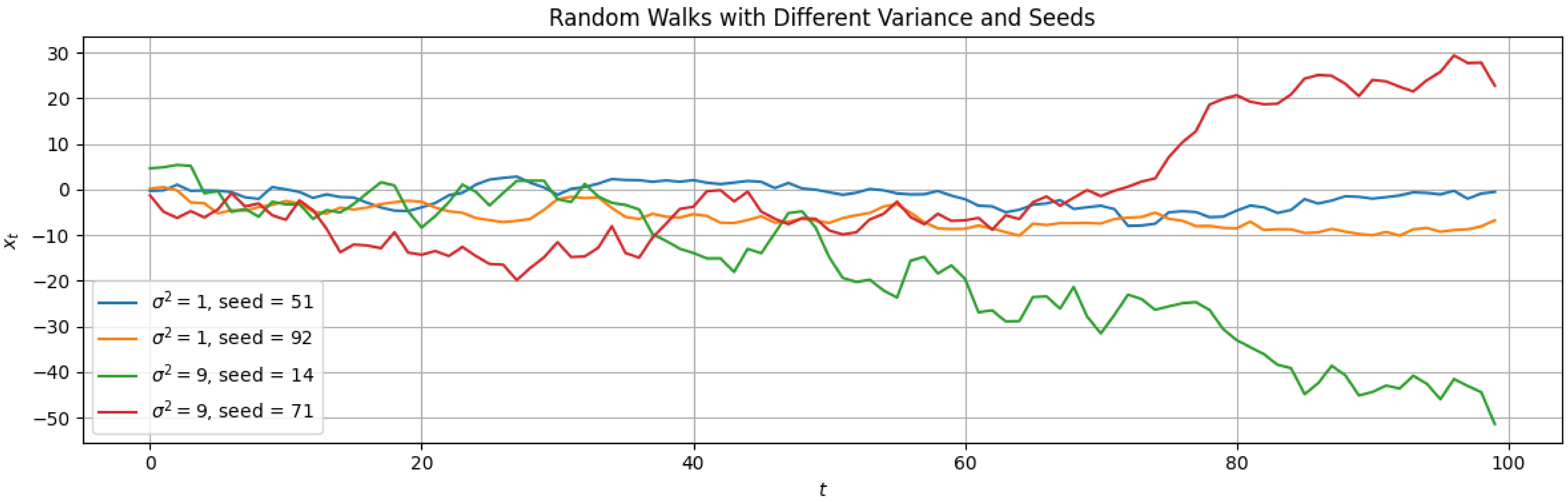
- Magnitude of trend and seasonal fluctuations are proportional to the value of the TS.
- Equivalent to the additive decomposition of the **log transform**ed TS.

Model with noise – Random walk (RW) process

naiv basline: repeat last value

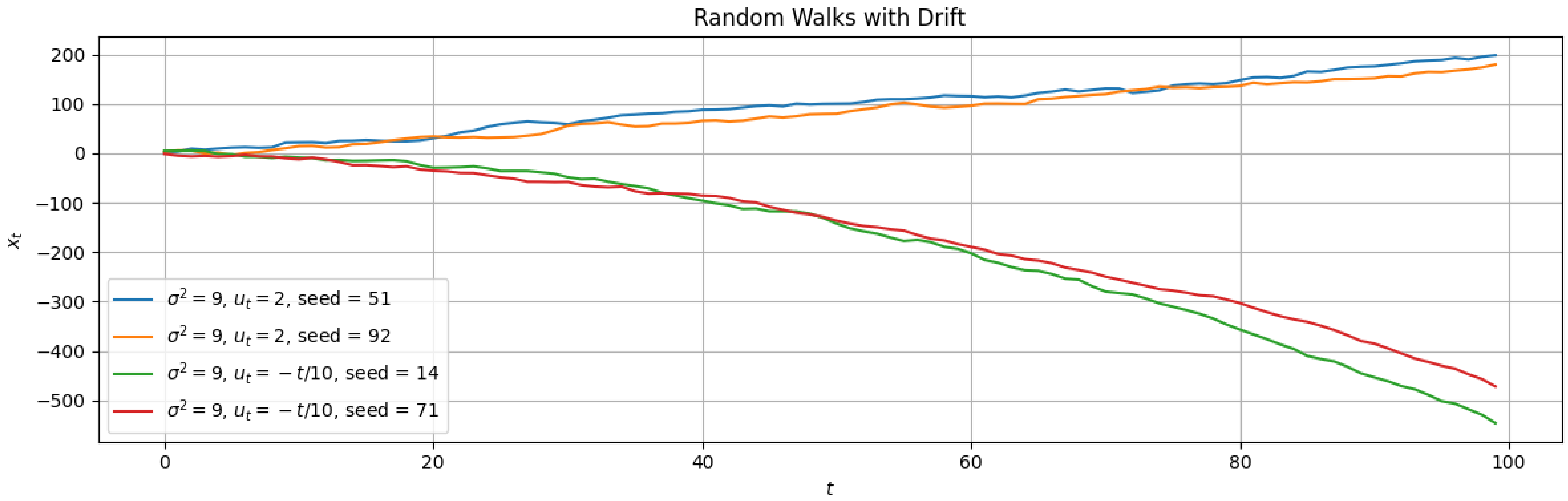
A **random walk** is a cumulative sum of iid zero-mean RVs: $X_t = \underbrace{0}_{U_t} + \underbrace{0}_{S_t} + \underbrace{\sum_{i=0}^{t} W_i}_{R_t}$ with $W_i \sim iid\ WN(0, \sigma^2)$

A RW process is **not iid** since its RVs are correlated: $X_t = X_{t-1} + W_t$ with $X_0 = W_0$.



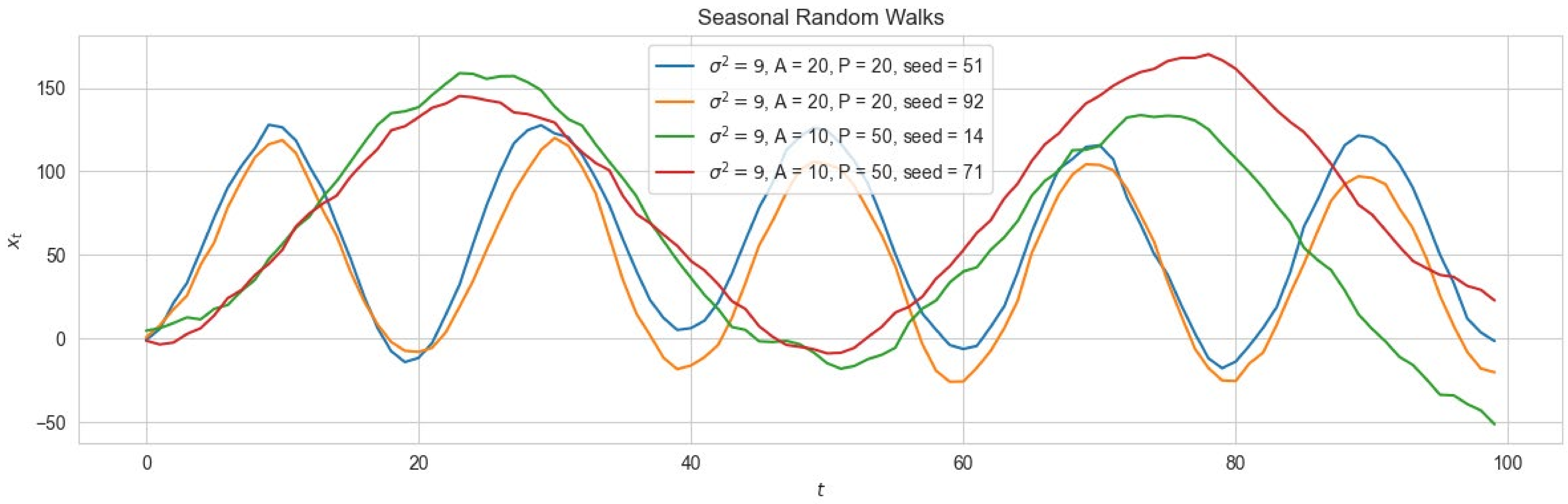
Model with trend – Random walk with drift

Random walk with drift: $X_t = X_{t-1} + W_t + U_t = \underbrace{\sum_{i=0}^{t} \mathrm{U}_i}_{U_t} + \underbrace{0}_{S_t} + \underbrace{\sum_{i=0}^{t} W_i}_{R_t}$ with $X_0 = W_0 + U_0$.



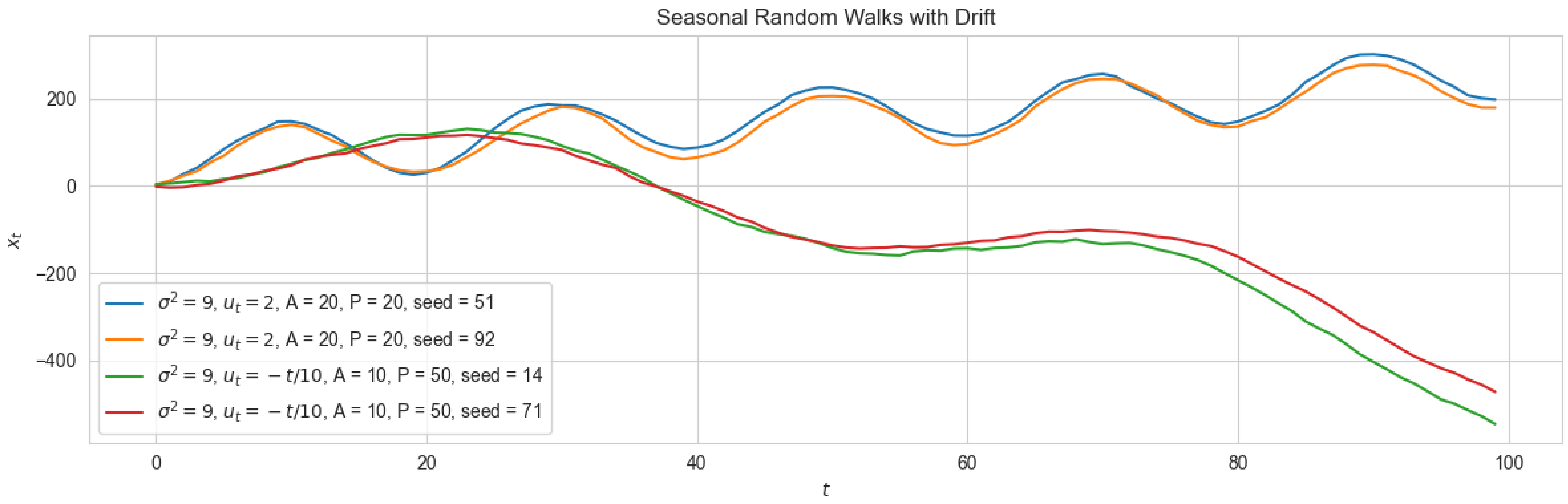
Model with seasonality – Seasonal random walk

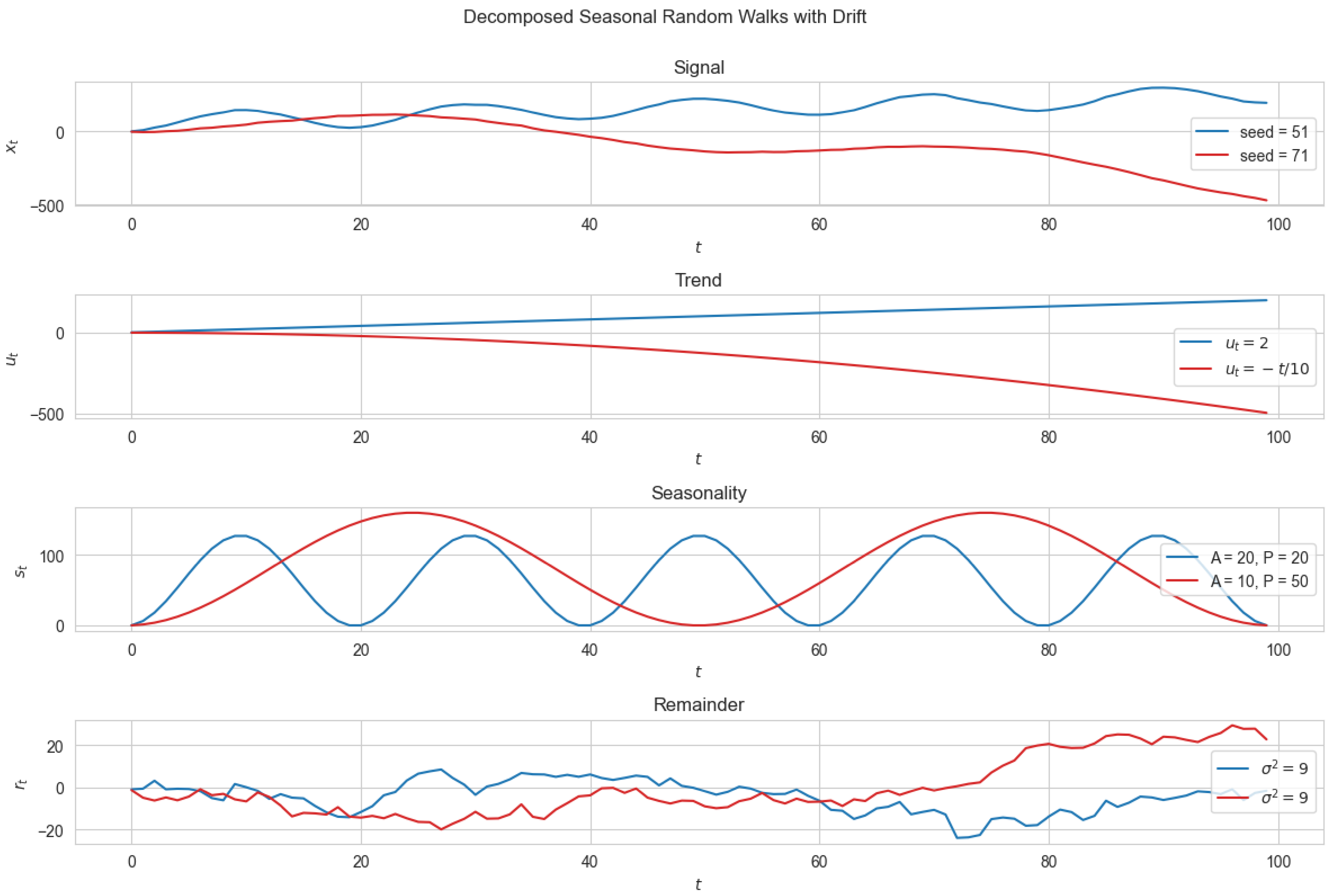
Seasonal random walk: $X_t = X_{t-1} + W_t + A\sin\left(\frac{2\pi t}{P}\right) = \underbrace{0}_{U_t} + \underbrace{\sum_{i=0}^{t}\left(A\sin\left(\frac{2\pi t}{P}\right)\right)}_{S_t} + \underbrace{\sum_{i=0}^{t} W_i}_{R_t}$ with $X_0 = W_0 + S_0$.



Model with trend and seasonality – Seasonal random walk with drift

Seasonal RW with drift: $X_t = X_{t-1} + W_t + U_t + A\sin\left(\frac{2\pi t}{P}\right) = \underbrace{\sum_{i=0}^{t} \mathrm{U}_i}_{U_t} + \underbrace{\sum_{i=0}^{t}\left(A\sin\left(\frac{2\pi t}{P}\right)\right)}_{S_t} + \underbrace{\sum_{i=0}^{t} W_i}_{R_t}$ with $X_0 = W_0 + U_0 + S_0$.




Decomposed Seasonal Random Walks with Drift
Signal
x_t
0
−500
0
20
40
60
80
100
t
seed = 51
seed = 71
Trend
u_t
0
−500
0
20
40
60
80
100
t
$u_t = 2$
$u_t = -t/10$
Seasonality
s_t
100
0
0
20
40
60
80
100
t
A = 20, P = 20
A = 10, P = 50
Remainder
r_t
20
0
−20
0
20
40
60
80
100
t
$\sigma^2 = 9$
$\sigma^2 = 9$

Decomposing time series with unknown components

Given a time series realization $\{x_1, x_2, \dots x_n\}$, assuming additive decomposition:

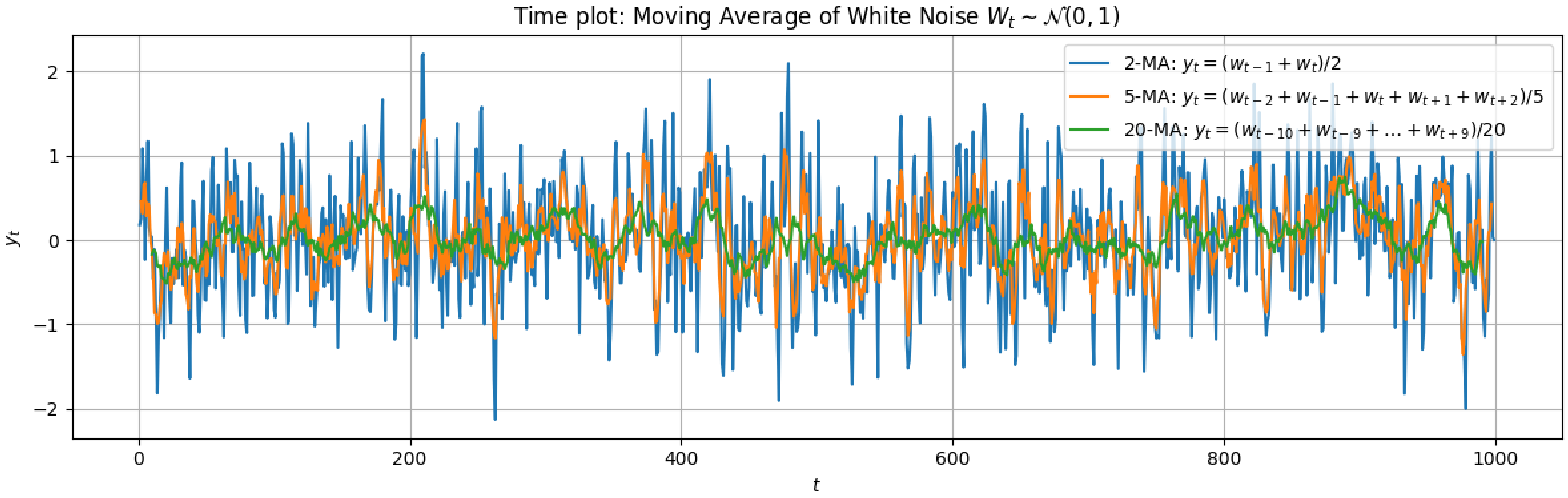
1. Estimate trend component $\hat{u}_t$
   – Moving average smoothing (classical decomposition)
   – LOESS – locally estimated scatterplot smoothing (STL decomposition)
2. **Detrend** time series $\hat{d}_t = x_t - \hat{u}_t$
3. Estimate seasonality component $\hat{s}_t$
   – Period adjusted averages (classical decomposition)
   – LOESS – locally estimated scatterplot smoothing (STL decomposition)
4. **Deseasonalize** TS to estimate the remainder $\hat{r}_t = \hat{d}_t - \hat{s}_t$

Estimating the trend – Moving average (MA) smoothing

Linear combinations of a TS values are called **linear filters**. The resulting TS is called a **filtered TS**.

A moving average of order m (m-MA) **smooths** a TS by averaging **m consecutive points**: $Y_t = \frac{1}{m}\sum_{i=-\lfloor(m-1)/2\rfloor}^{\lfloor(m-1)/2\rfloor} X_{t+i}$

(window of size m)



Estimating the trend – Moving average (MA) smoothing

Variations

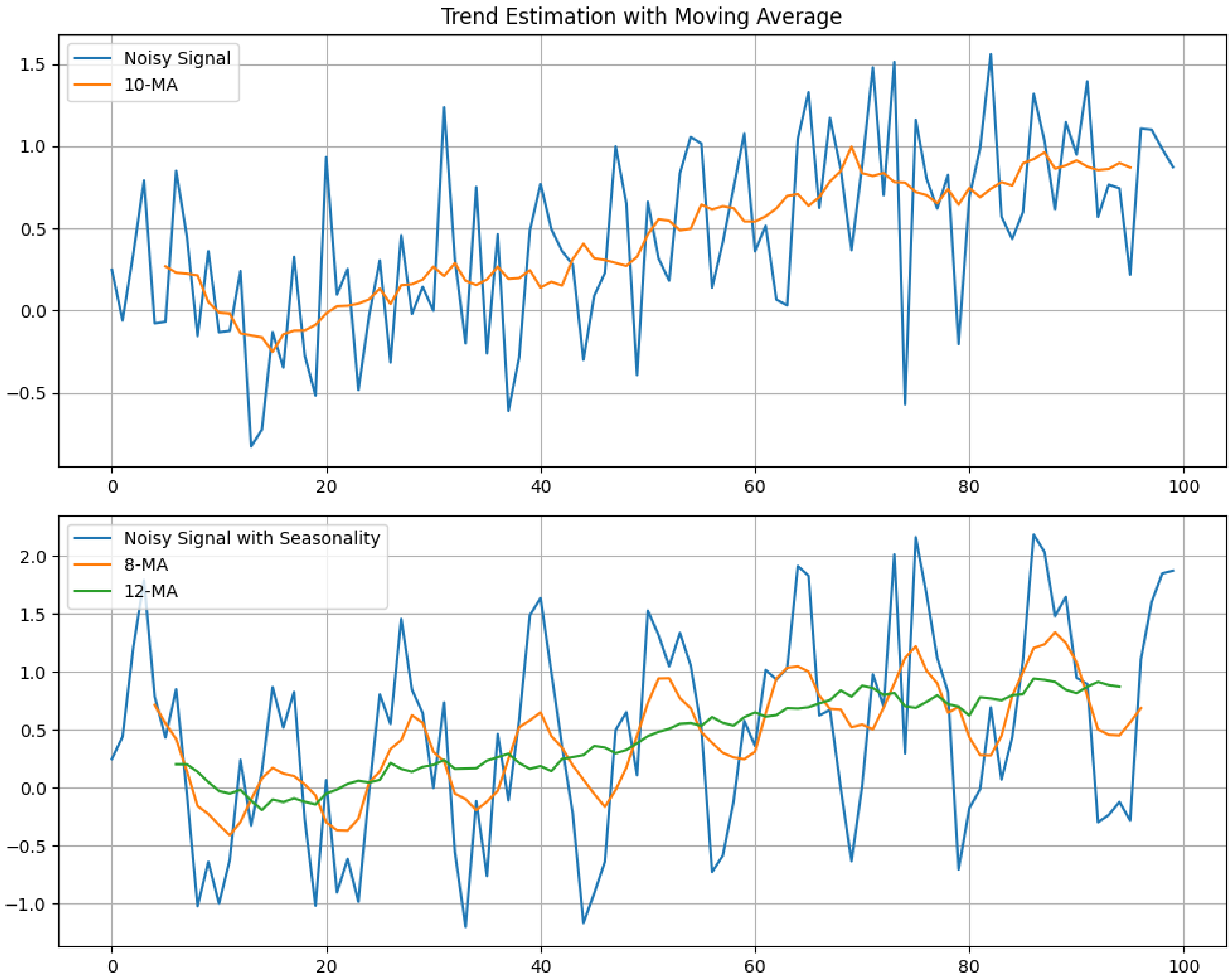
- **Window center can be shifted** e.g., trailing 3-MA $Y_t = \frac{1}{3}(X_{t-2} + X_{t-1} + X_t)$
- Different weights can be assigned observations in the window (**weighted MA**) $Y_t = \frac{1}{m}\sum_{i=-\lceil(m-1)/2\rceil}^{\lfloor(m-1)/2\rfloor} a_i X_{t+i}$

MA can be applied iteratively e.g., 4-MA then 2-MA referred as 2 x 4-MA.

- Additional smoothing
- Symmetry for even orders: 2 x m-MA is equivalent to a weighted (m+1)-MA with $w = \left[\frac{1}{2m}, \frac{1}{m}, \frac{1}{m} \dots, \frac{1}{m}, \frac{1}{2m}\right]$

MA for trend estimation $\hat{u}_t = y_t$:

- Smooth out fluctuations to reveal underlying trends and cycles.
- Remove seasonal components to understand trend and cyclical behavior: **match order with period**.
- **Sensitive to outliers**


Trend Estimation with Moving Average
Noisy Signal
10-MA
Noisy Signal with Seasonality
8-MA
12-MA

Estimating the seasonality – Period adjusted averages

Given a detrended TS realization $\{d_1, d_2, \dots d_n\}$ with period P, assuming additive decomposition:

1. **Group** seasonal values
   - For $t = 1,2, \dots P$, collect all detrended values d_{t+iP} that fall at position t in each cycle.
   - For example, with monthly data and yearly period, group all Jan values, Feb values, etc.

2. **Average** within each group to get the raw seasonal estimates.

3. Adjust the components so they **sum to zero** (since we are considering the detrended TS)
   - Subtract the overall average of the seasonal estimates.

4. **Repeat** the seasonal component values for each period
   - Assign each time point the seasonal value of its cycle position.
   - For example, Jan gets the Jan seasonal, Feb the Feb seasonal, etc.

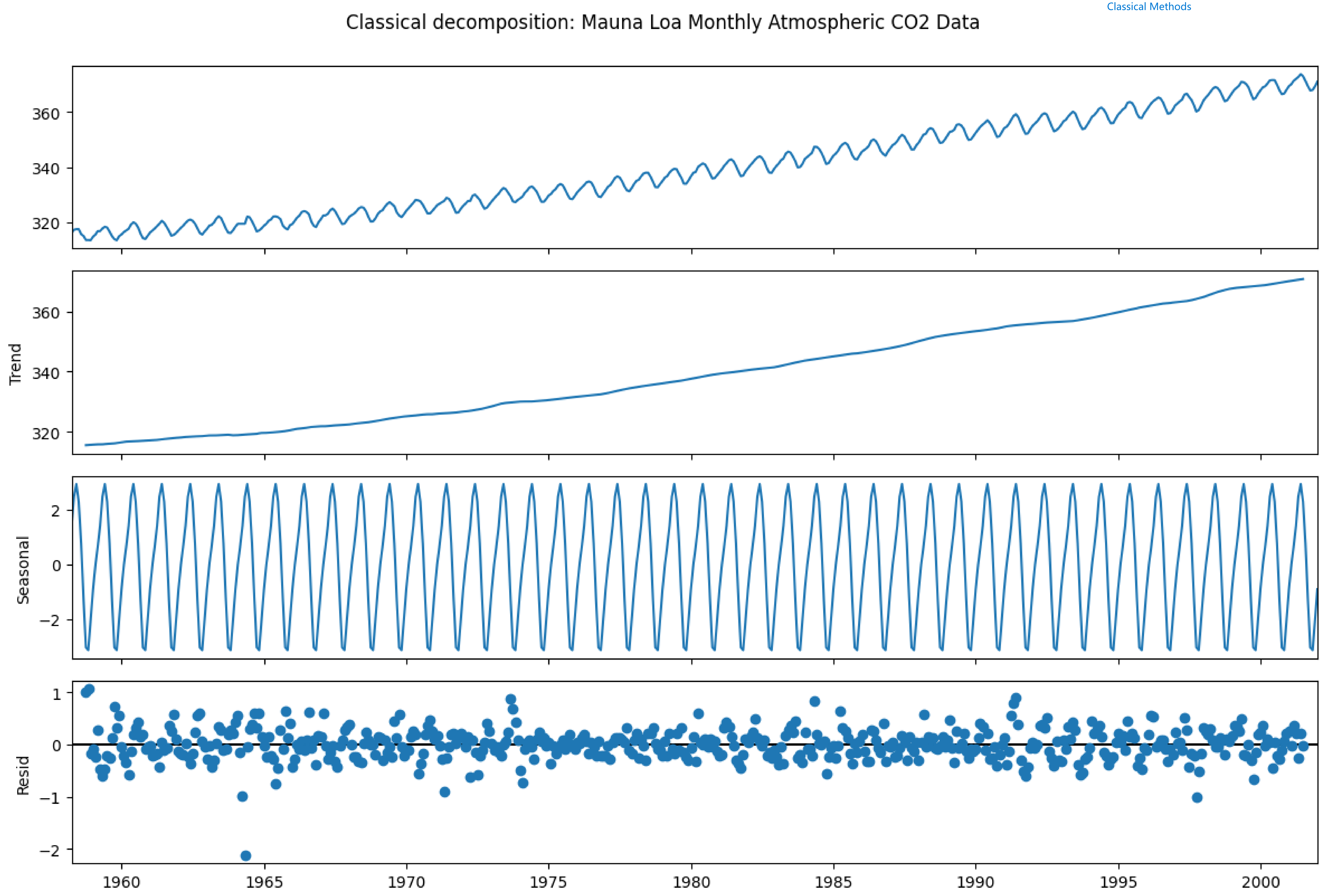
Seasonal and trend decomposition using LOESS (STL)

LOESS (locally estimated scatterplot smoothing) is a non-parametric regression method used to fit data.

- **Generalization** of moving average and polynomial regression able to **capture non-linear trends**.
- Apply a **sliding window** across the dataset, where at each point, a small subset of neighboring points is selected to **fit a local linear/polynomial model**.
- Neighbors are weighted according to their proximity.
- The size of the window (span) determines the **smoothing strength**.

STL uses LOESS to estimate both the trend and seasonality components

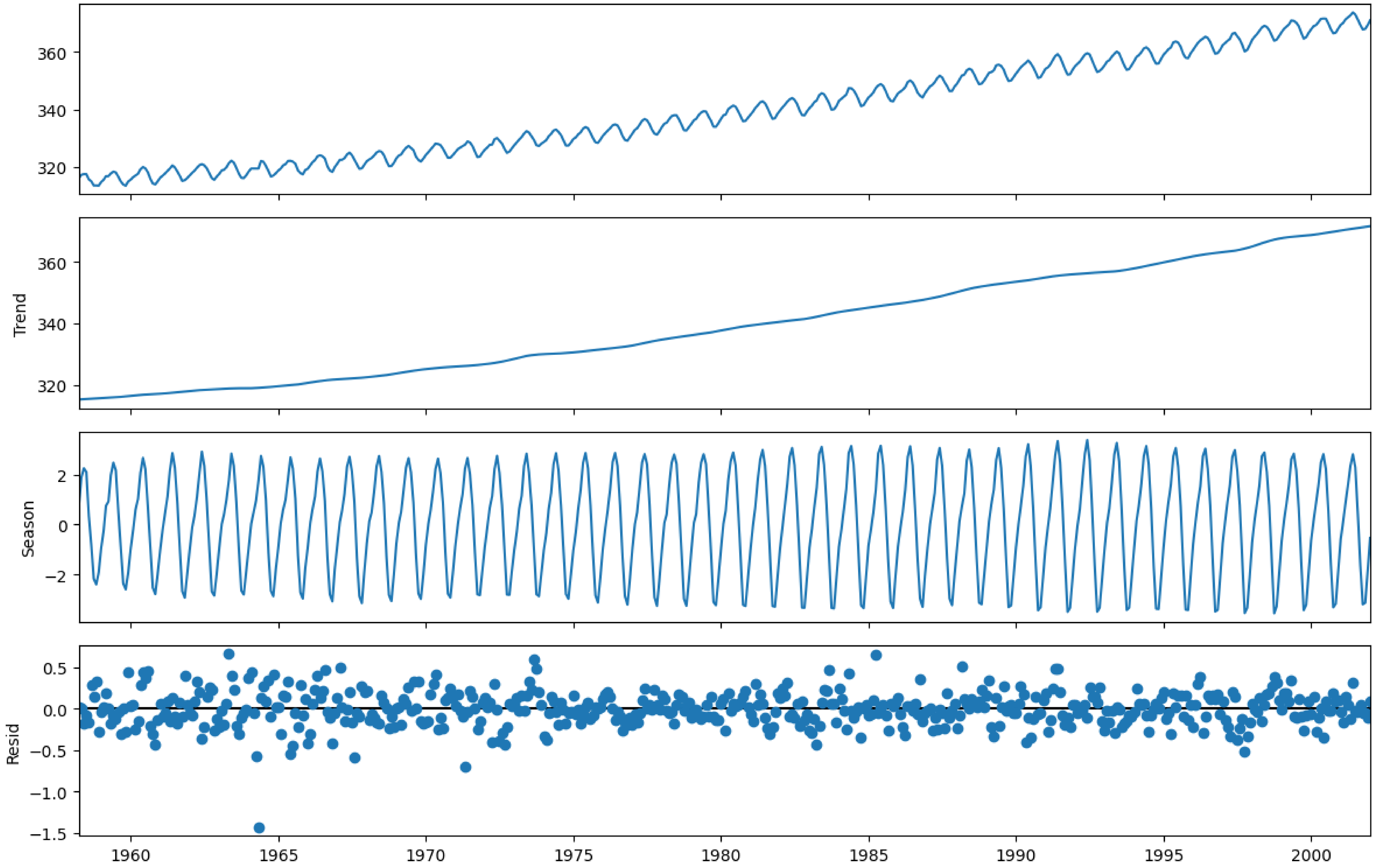
- Different sliding window spans for trend and seasonality
- In comparison to period adjusted averages, **seasonal component is allowed to change over time**.
- Can down-weight outliers to reduce their impact.



statsmodels.tsa.seasonal.seasonal_decompose

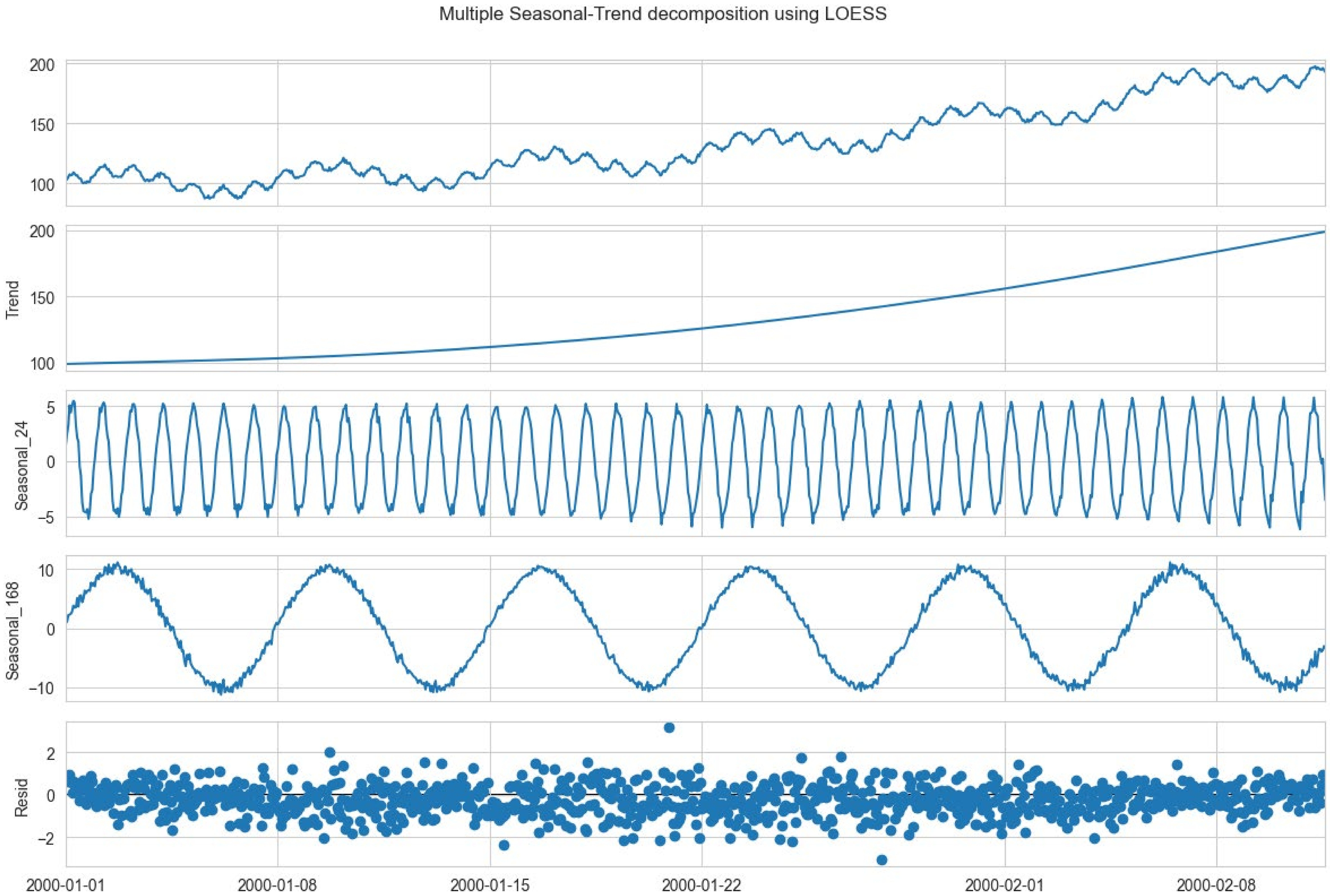
Keeling et al. "Atmospheric CO2 concentrations derived from flask air samples at sites in the SIO network." Trends: a compendium of data on Global Change (2004).

STL decomposition: Mauna Loa Monthly Atmospheric CO2 Data



statsmodels.tsa.seasonal.STL

Keeling et al. "Atmospheric CO2 concentrations derived from flask air samples at sites in the SIO network." Trends: a compendium of data on Global Change (2004).



statsmodels.tsa.seasonal.MSTL

Exercise

Generate 2-3 **synthetic time series** with known components.

- Apply classical and STL time series decomposition.
- Review how well components are extracted, compute the **mean squared error**.

Generate and **interpret** the ACF plots of the different forms of random walks presented in this lecture.

Reimplement classical decomposition.

Extend lecture 1 exercise with ACF plots and time series decomposition.

- **Compare** classical and STL decomposition results.
- **Interpret** ACF plots and time series components.