

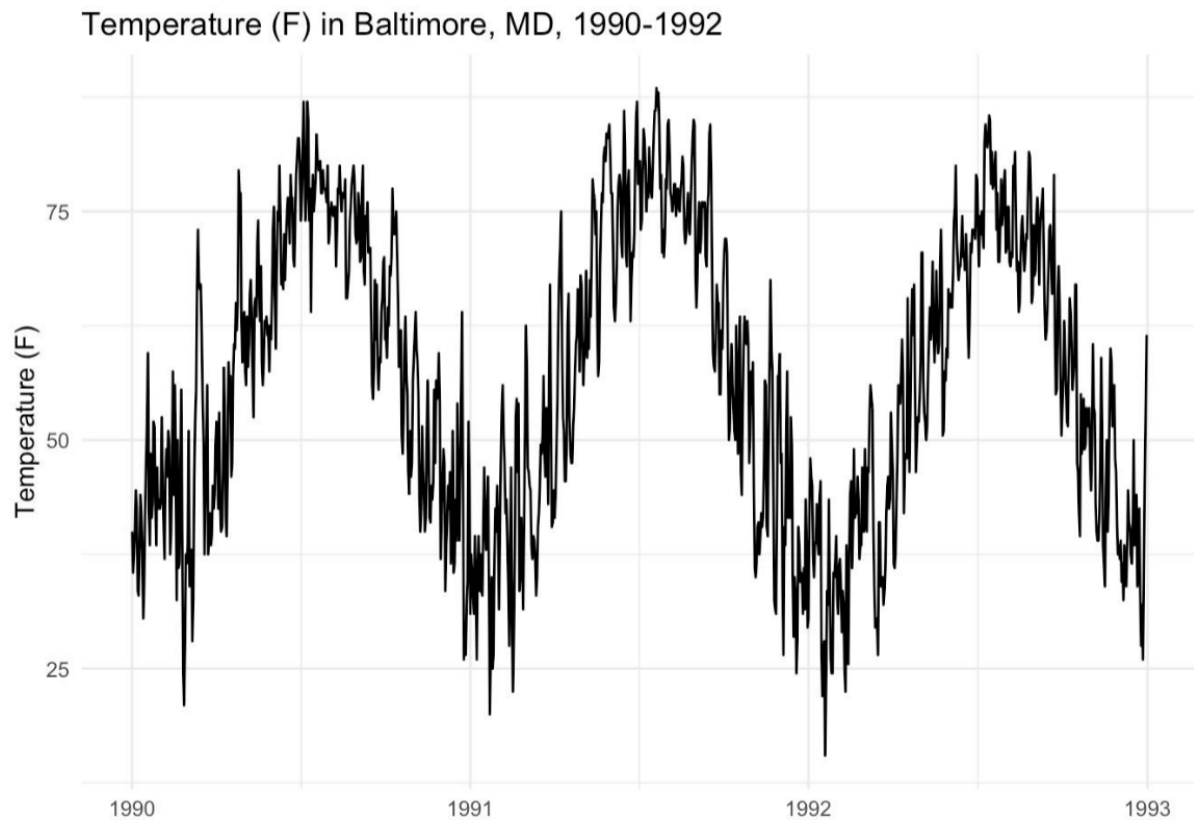
# Time Series Data

MIE223  
Winter 2025

## 1 Time Series Analysis

### 1.1 Time Series Data

- Time series data are
- Observations or measurements that are indexed according to time
- Instead of  $X_i$  we have  $X_t$  where  $t$  is specifically a time index
- Why does time make it different?
- The time index  $t$  has a special ordering
- Data measured over time are not exchangeable, which is what we often assume when data are indexed by  $i$



## 1.2 Goals of Time Series Analysis

- Time Scale Analysis: at what scale is data useful to analyze?
- Is weather useful to analyze at the second, hour, or year level? Is it seasonal?
- Smoothing: infer true state of past given a noisy signal (averaging out nearby timepoints)
- Assuming gradual change, use nearby values in a time series to smooth noise
- Filtering: what is the true present state given a noisy signal
- Tracking a rocket trajectory
- Kalman filtering!
- Forecasting: predict future values
- What will the stock price be tomorrow, given the past?
- Regression: give time two series, can we predict the association between them?
- Does public sentiment predict the stock market at a delay?

## 2 Time Scale Analysis

### 2.1 Time Scale Analysis

- What time scales of variation dominate or account for the majority of temporal variation in the data? seasonal
- Is there a strong seasonal cycle in the observations of temperature in Baltimore, MD? Hope so
- Is the association between ambient air pollution and mortality in a city primarily driven by large annual changes in pollution levels or by short-term spikes? there can be multiple causes of mortality, and need longer timeframe
- granger causality says that you can predict something if you know the past

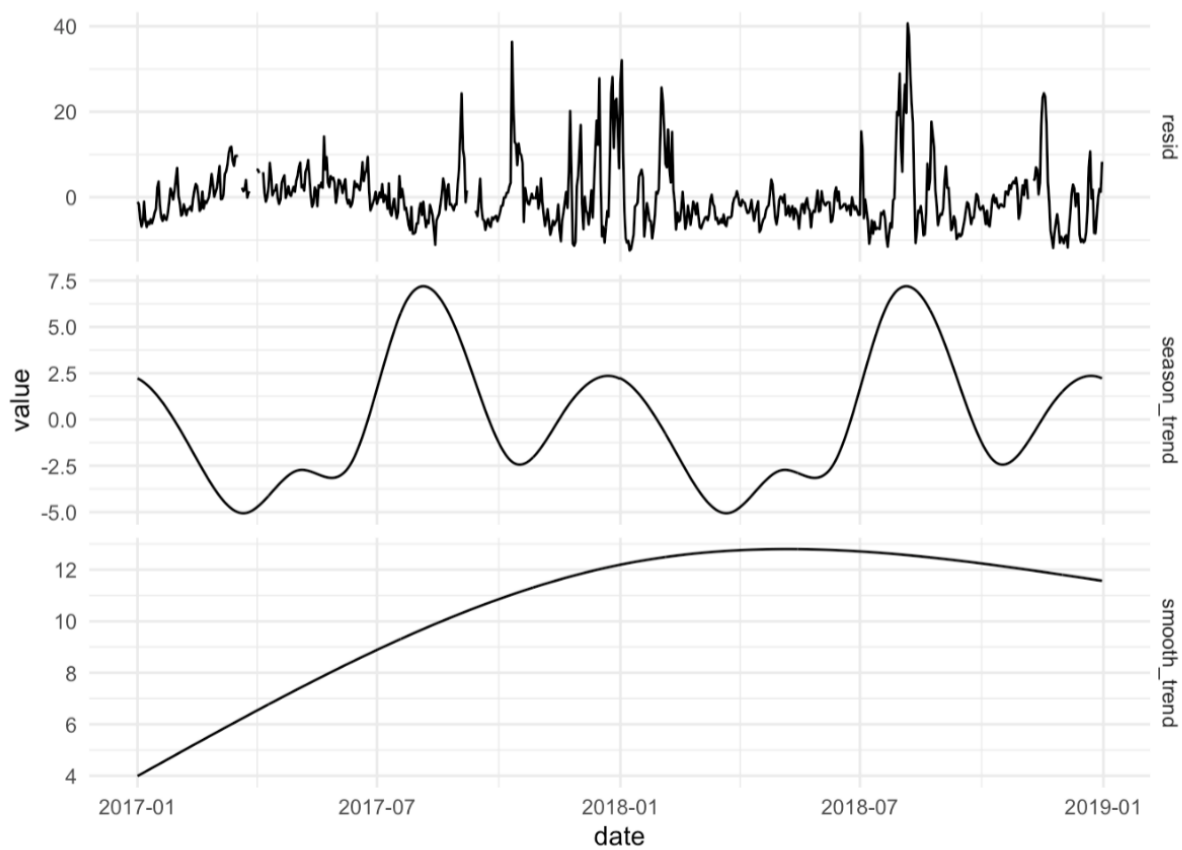
### 2.2 Trend-Season-Residual Decomposition

A common tool to use is to decompose a time series into

1. Smooth long-term trend (trend)
2. Seasonal variation (seasonal)
3. Residual variation (day to day)

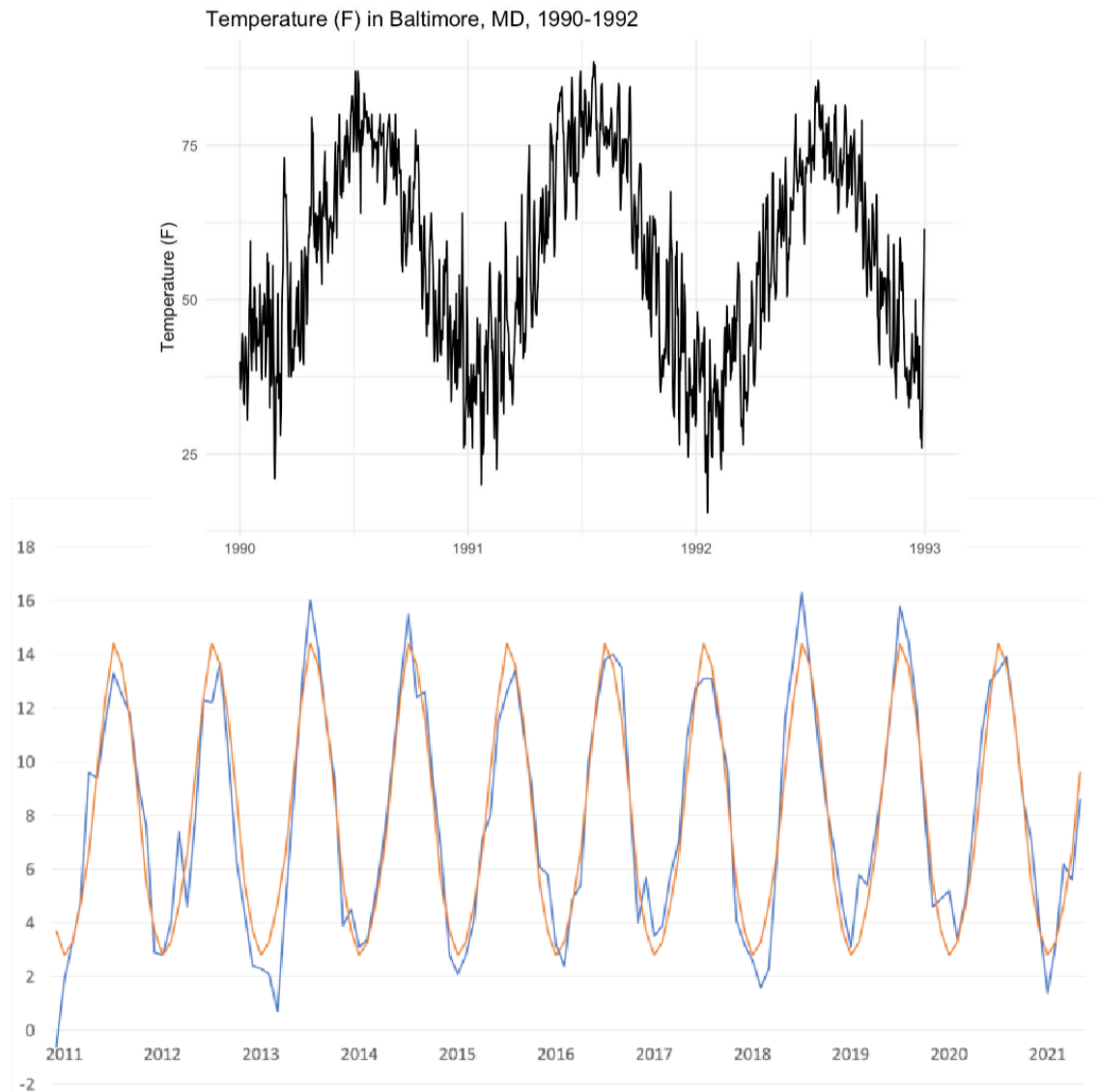
The primary benefit is of looking at long-term trend and seasonal variation is that they are highly interpretable and are very common.

These graphs can be made by smoothing over different intervals.



### 2.3 Trend-Season-Residual Example 1

- Daily and monthly temperature vs. year • What time scales are informative?
- What do you think could be
  - Long-term trend: climate change
  - Is there seasonality? weather seasons
  - Cause of variation? day to day fluctuations in the weather



## 2.4 Trend-Season-Residual Example 2

Three stocks: GOOG, MSFT, ZOOM • Which is which?

- What do you think could be
  - Long-term trend
  - Is there seasonality?
  - Cause of variation?

## ual Example 2



## 3 Time Series Regression with Autocorrelation and Cross-correlation

### 3.1 Regression (in general)

Given a time series of two phenomena, what is the association between them?

- What is the association between daily levels of air pollution and daily values of cardiac hospitalizations?
- What is the lag (in months) between a change in a country's unemployment rate and a change in the gross domestic product?
- What is the cumulative number of excess deaths that occurs in the 2 weeks following a major hurricane?

### 3.2 Recap of Covariance

- Covariance recap:
  - Two random variables  $X, Y$  with means  $E(X), E(Y)$
  - Covariance  $Cov_{X,Y} = E((X - E(X)) \cdot (Y - E(Y)))$
  - If  $X$  and  $Y$  are independent:
    - $Cov_{X,Y} = E(XY - E(X)Y - E(Y)X + E(X)E(Y)) = 0$
  - If  $Cov_{X,Y} > 0$ :
    - High values for  $x$  go with high values for  $y$
  - If  $Cov_{X,Y} < 0$ :
    - High values for  $x$  go with low values for  $y$

### 3.3 Pearson Correlation Coefficient

Covariance depends on the units in which  $X$  and  $Y$  are given

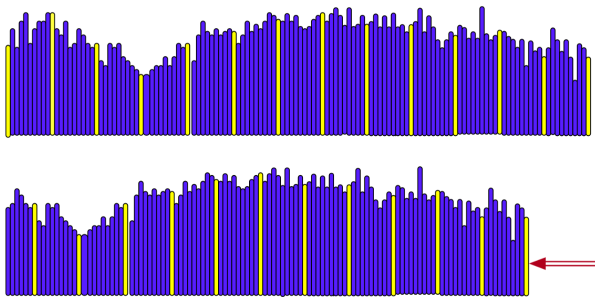
Therefore, we look at the Pearson correlation coefficient

If  $x$  and  $y$  have standard deviations  $\sigma_x$  and  $\sigma_y$  respectively:

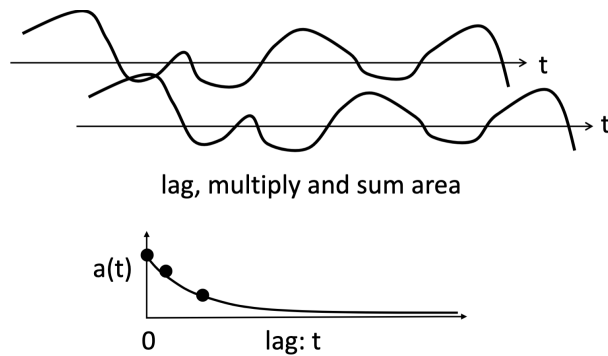
$$\begin{aligned} \rho_{X,Y} &= \frac{Cov_{X,Y}}{\sigma_X \sigma_Y} \in [-1, 1] \\ &= \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{\sqrt{n \sum x_i^2 - (\sum x_i)^2} \sqrt{n \sum y_i^2 - (\sum y_i)^2}} \end{aligned}$$

### 3.4 Autocorrelation

Auto-correlation: Compare the correlation of a time series with itself moved (lagged) by  $k$  positions



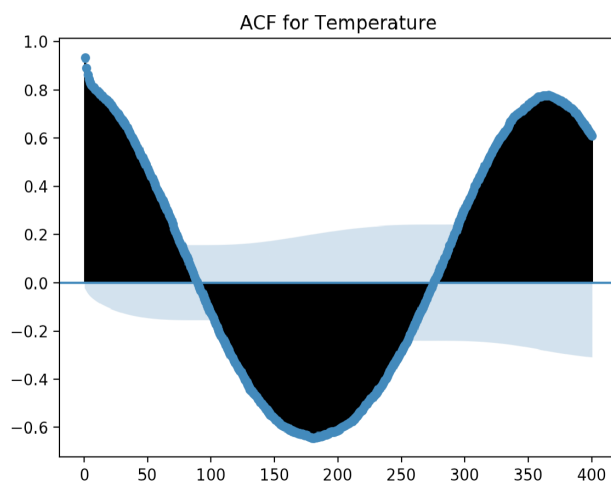
### 3.5 Autocorrelation: Measure of correlation in time series at different lags



### 3.6 Autocorrelation Example

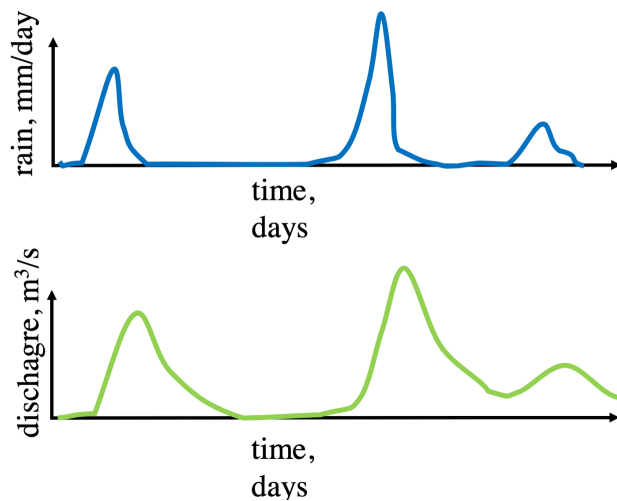
Autocorrelation for Chennai temperatures

- High auto-correlation for the next day and for a year afterwards
- The light blue denotes significance



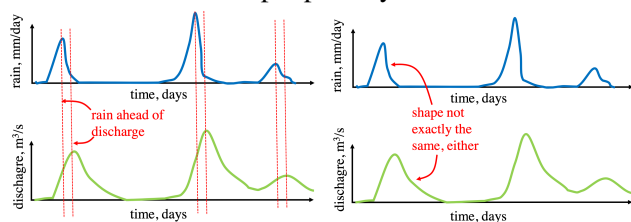
### 3.7 Cross-correlation

- Consider how one time series “predicts another”
  - Example: how does the amount of rainfall affect (predict) the discharge water flow rate of a river measured at a given point?
    - \* Discharge correlated with rain
    - \* Discharge is delayed behind rain because rain takes time to drain from the land

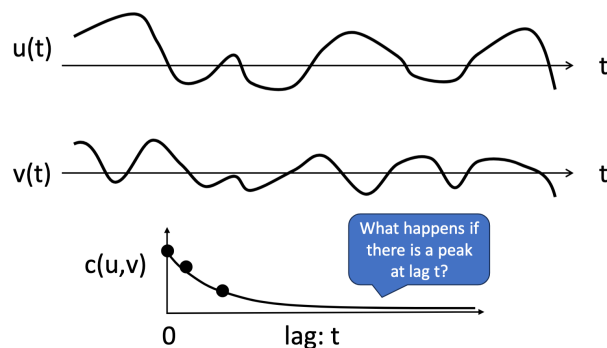


Sample midterm question: Would you use VADER to predict the stock market?

Answer: No, VADER is a sentiment analysis tool that is used to analyze text data. It is not used to predict the stock market. Simple polarity sentiment is not complex enough to analyze the stock market



### 3.8 Cross-correlation Measure of correlation between two time series at different lags



## 4 Smoothing