


TIME SERIES ANALYSIS AND APPLICATIONS

Prof. Luana Medeiros Marangon Lima, Ph.D.

Durham, June 15th 2023

Energy Data Analytics PhD Student Fellows Workshop

Let’s start with introductions…

- Originally from Brazil
- B.Sc. & MS. , Electrical Engineering

  FEDERAL UNIVERSITY OF ITAJUBÁ, Brazil, 2005 & 2007

  Electrical Power Systems
- Ph.D., Operations Research and Industrial Engineering

  UNIVERSITY OF TEXAS, Austin, TX, 2011

  Modeling and Forecast of Brazilian Reservoir Inflows via Dynamic Linear Models under Climate Change Scenarios
- Joined the Nicholas School in Jan 2018 and the EI/NIEES in July 2019
- Recent courses: Time Series Analysis for Energy Data, Environmental Data Analytics, Modeling for Energy Systems, Economics of Modern Power Systems
- Research and work experience: Renewable energy development and integration, Co-Optimization of Water & Energy Systems, Power System Economics, Climate Change effect on energy production

Agenda

- Intro to Time Series Analysis
  - Background and motivation
  - TS Definition and Components
  - Examples of TS Models
  - Applications to Energy
- Time Series Decomposition
  - Trend component and seasonal component
  - How to identify/estimate components
  - Case Study
- Forecasting

Intro to Time Series Analysis

Background and Motivation

TS Models

What is a Time Series (TS)?

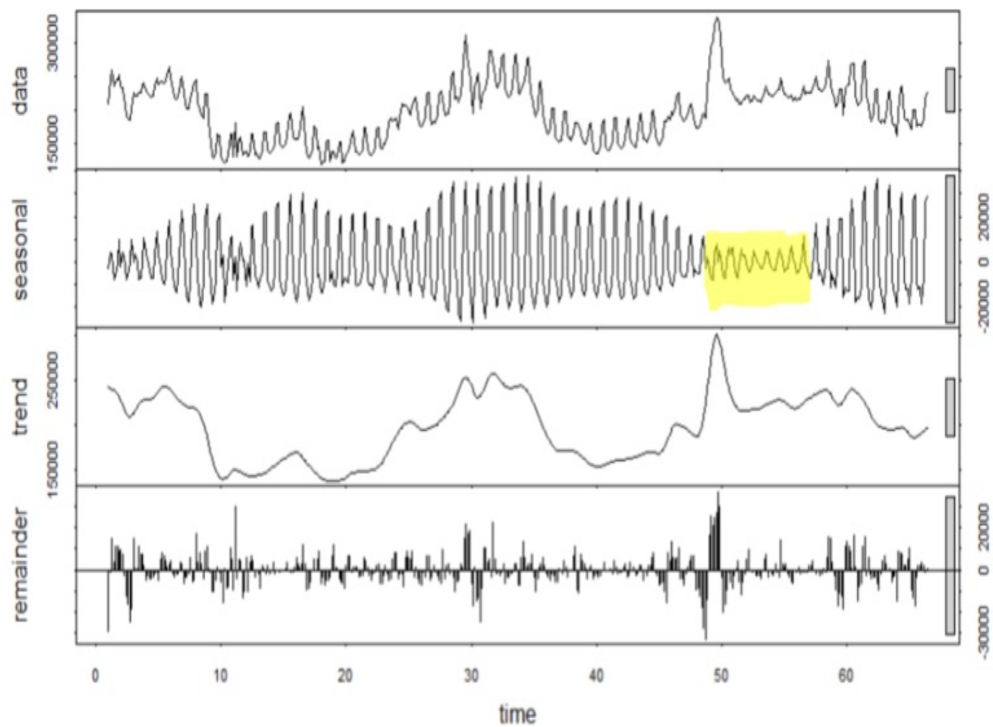
- Sequence of observations on a variable collected over time
  - Ex: stock prices, interest rate, retail sales, electric power consumption
- Mathematical representation
  - TS is defined by the values $Y_1, Y_2, \ldots$ of a variable Y at times $t_1, t_2, \ldots$ Thus,

$$Y = F(t)$$

Causes of Variation in TS data

- Calendar: seasons, holidays, weekends
- Example

Video game usage over time – daily basis



Interest

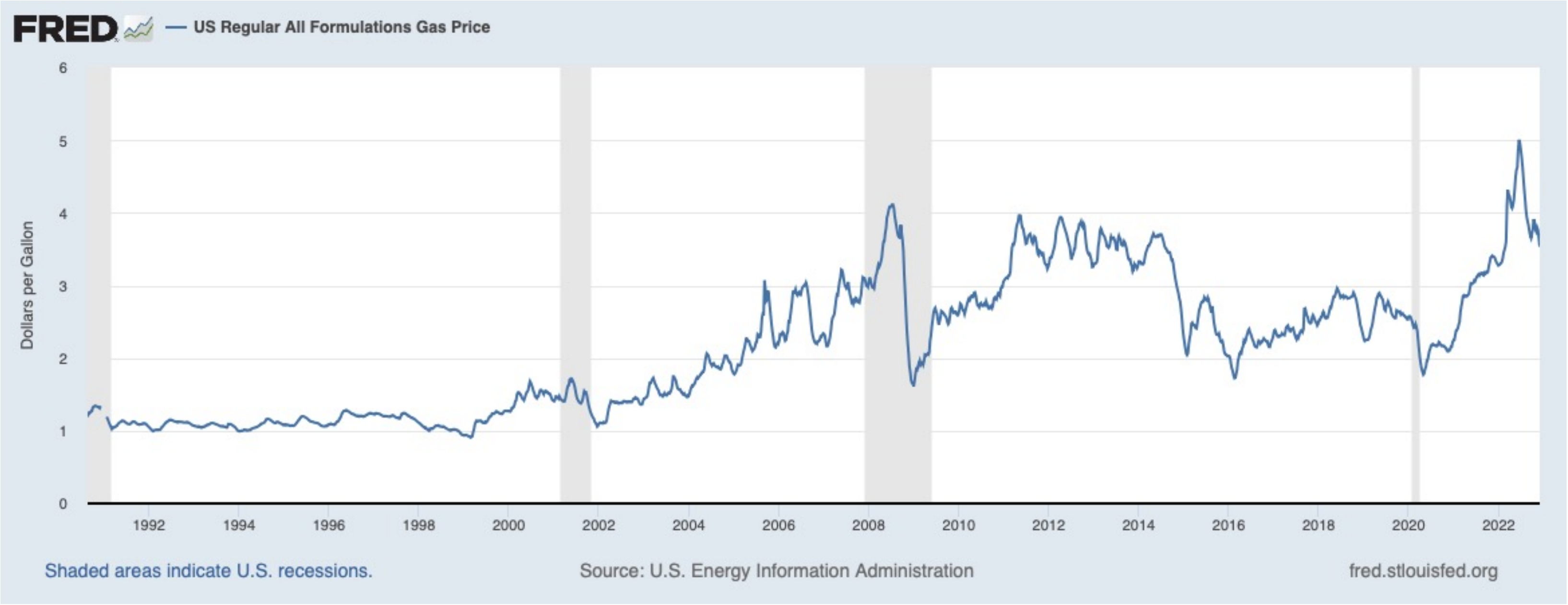
- Trend of usage over time
- Popularity of games
- Weekly cycle of usage

Knowledge

- Usage up on the weekend and down during the week
- Increased usage during holidays
- Summer time: weekdays/weekends blend together

Causes of Variation in TS data

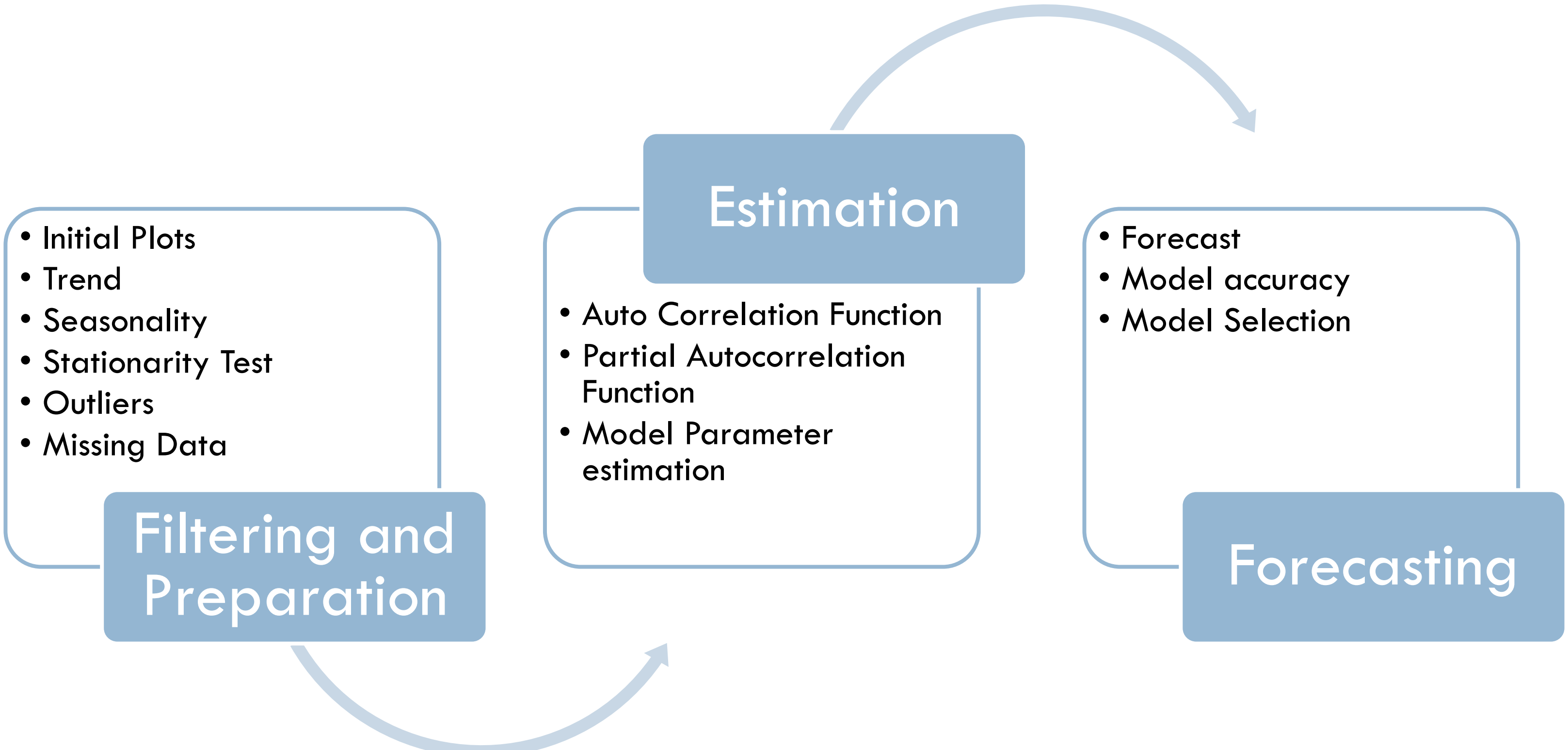
- Natural calamities: earthquake, epidemic, flood, drought
- Human: Political movements or changes, policies, war
- Example



Source: https://fred.stlouisfed.org/series/GASREGW#

What is Time Series Analysis (TSA)?

- In TSA, we analyze the past behavior of a variable in order to predict its future behavior



Why Learn Time Series Analysis?

Understanding **temporal patterns** and **dependencies** in the data →

- Forecasting and Prediction
- Resource Optimization
- Anomaly Detection and Intervention
- Policy Formulation and Evaluation
- Data-Driven Decision Making
- Risk Management

- Critical skill in today's **data-driven world**
- Widely applicable across various industries and domains

Examples of Time Series Models

- Autoregressive Integrated Moving Average (ARIMA)
  - combines three components: autoregression (AR), differencing (I), and moving average (MA)
- Seasonal Autoregressive Integrated Moving Average (SARIMA)
  - extension of ARIMA that incorporates seasonal components
- Exponential Smoothing Models: Exponential
  - Simple Exponential Smoothing (SES), Holt's Linear Exponential Smoothing, and Holt-Winters' Seasonal Exponential Smoothing
- Seasonal Decomposition of Time Series (STL): STL is a method that decomposes a time series into seasonal, trend, and residual components

TSA Applications to Energy

- **Energy demand** forecasting for resource planning, load balancing, and infrastructure development
- **Renewable energy** forecasting to facilitate solar and wind power grid integration
- Forecasting **energy prices** to inform trading strategies and risk management
- **Energy storage** optimal sizing and dispatch strategies for energy storage systems
- Identifying **abnormal energy consumption or production** patterns that may indicate equipment malfunction or faults
- **Load profiling** for dynamic pricing structures and personalized tariffs/rates to incentivize energy efficiency

Time Series Decomposition

Trend Component

Seasonal Component

Time Series Components

- A time series may have the following components:

Trend Component

Seasonal Component

Cyclical Component

Random Component

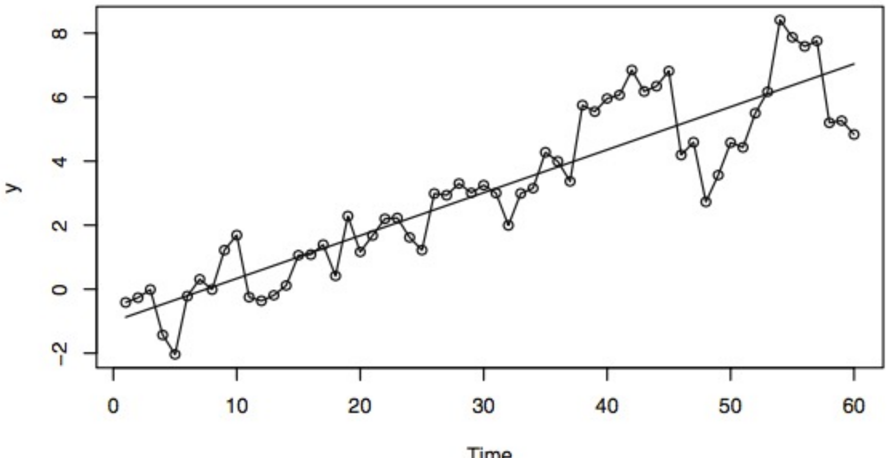
Decomposing the Time Series means separating trend/cycle, seasonal and random components.

TSA will find and exploit predictable patterns/components.

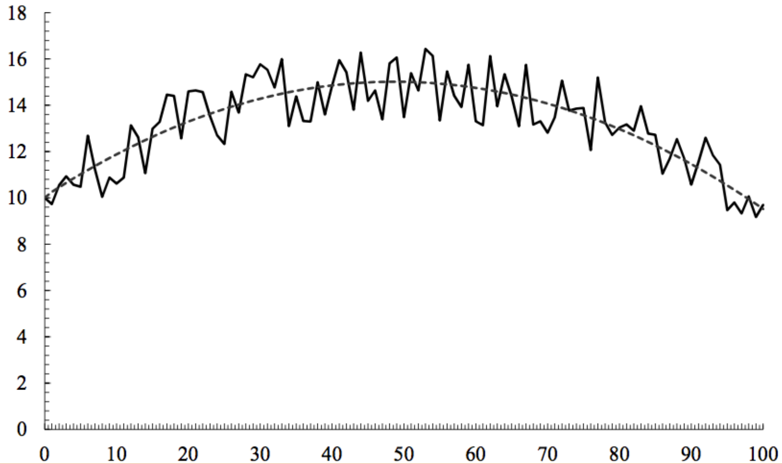
Trend Component

- Long-term tendency
  - Increase (upward movement) or
  - Decrease (downward movement)
- Trend can be linear or non-linear

Ex: Upward Linear Trend



Ex: Quadratic Trend



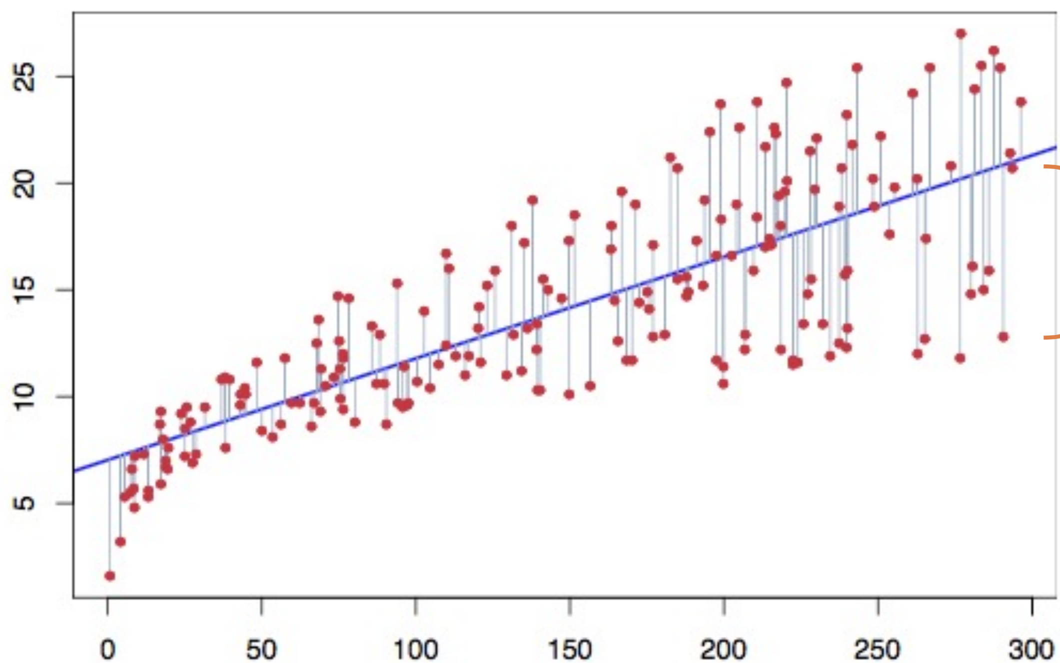
Most of the time we assume a rolling average to simplify the analysis

Linear Trend Component

- For a linear trend we can write

$$Y_i = \beta_0 + \beta_1 t_i + \varepsilon_i$$

- **Slope** (β_1) and the **intercept** (β_0) are the unknown parameters, and ε_i is the **error term**



$\widehat{Y}_i = \beta_0 + \beta_1 t_i$

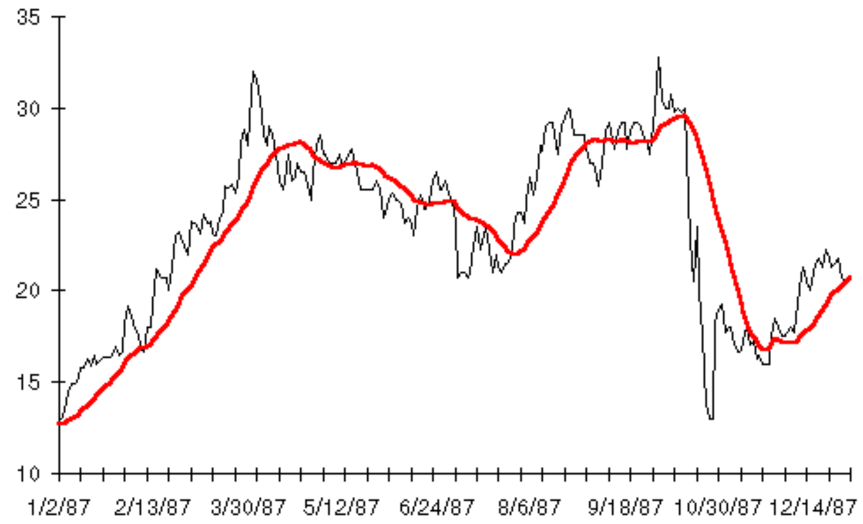
The error term or residual is the distance from point Y_i to the estimate $\widehat{Y}_i$

$$\varepsilon_i = Y_i - \widehat{Y}_i$$

Non-Linear Trend Estimation: Moving Average

- Smooth out the trend with something like a rolling average
  - A moving average trendline smooth out fluctuations in data to show a pattern or trend more clearly
  - Which order to use for the moving average?

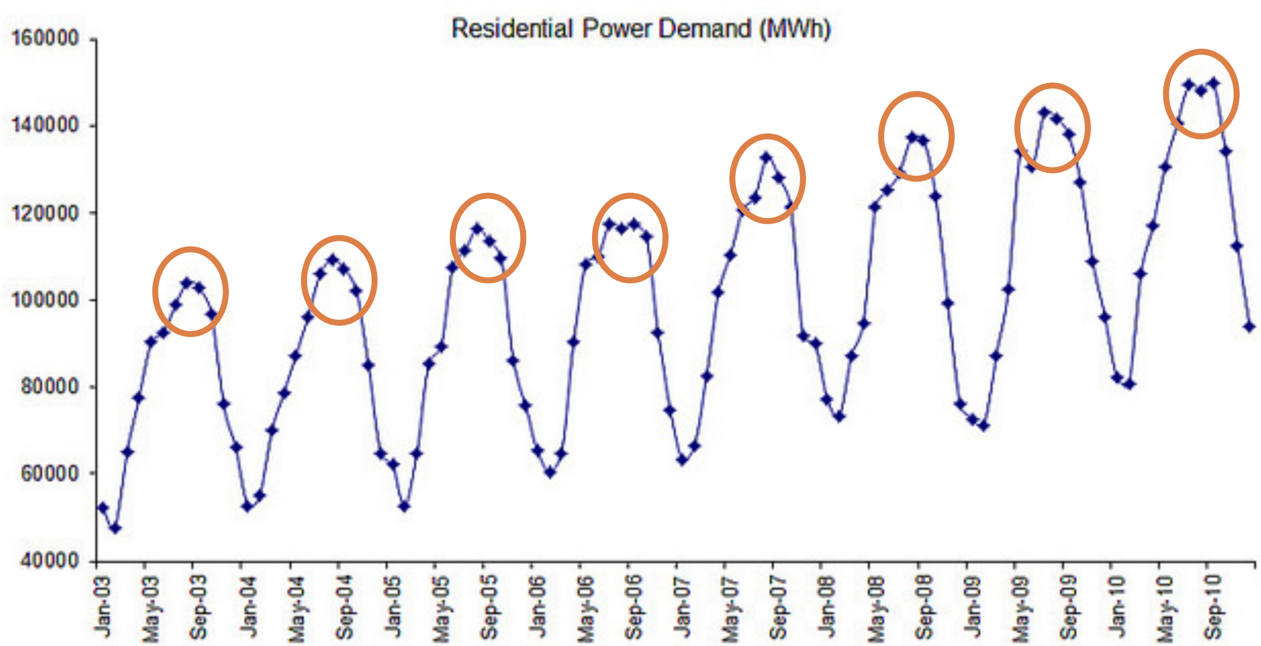
- Looking at the rolling average makes it easier to tell how the trend is moving underneath the noise



Seasonal Component

- Short-term regular wave-like patterns
  - Observed within 1 year - monthly or quarterly
  - Equally spaced peaks and troughs

Calendar Related



Peaks in the Summer months Jun/Jul/Ago

Decomposition: Trend Component

1. **Smooth the trend** with a moving average
   - Find Y_{trend}

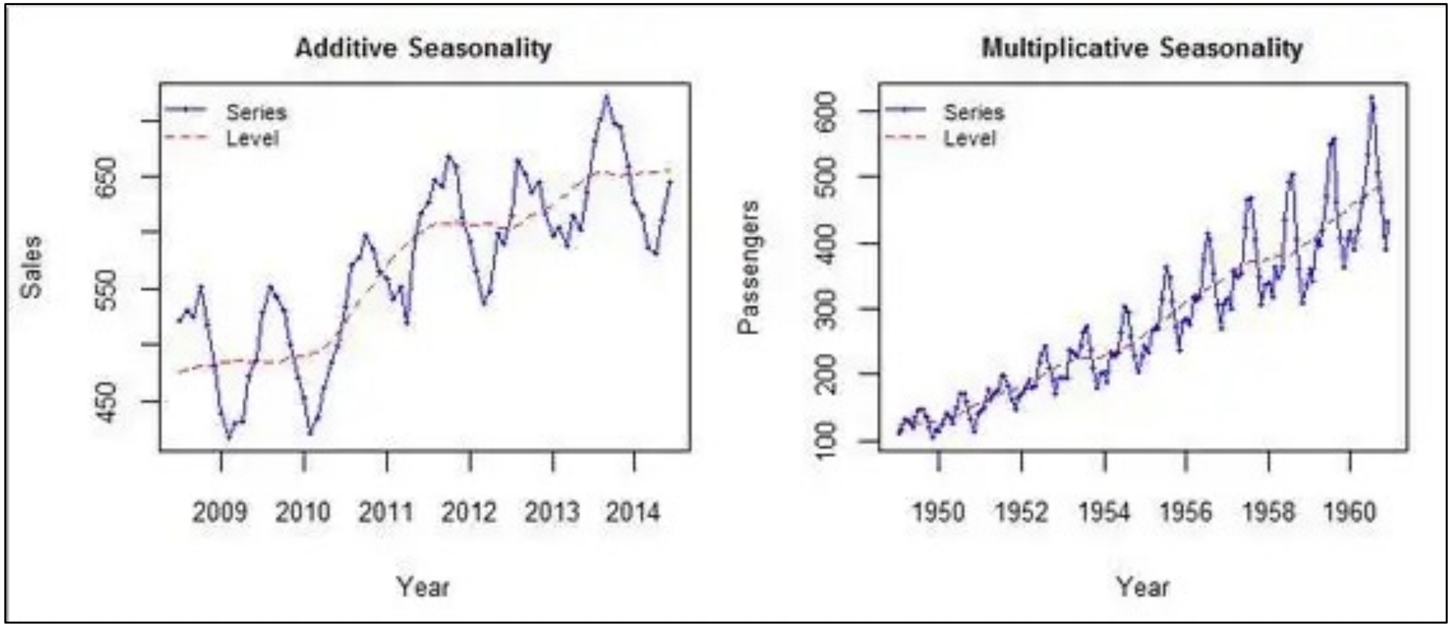
2. **De-trend** the series
   - Additive Model
     - take original series and **subtract** the smoothed trend

$$Y_{seasonal} = Y - Y_{trend}$$

   - Multiplicative model
     - scales the size of the seasonal component as the trend rises or falls
     - take original series and **divide** the original data by the trend

$$Y_{seasonal} = \frac{Y}{Y_{trend}}$$

Additive vs Multiplicative Model



- In the **additive model** the magnitude of seasonality does not change in relation to time
- In the **multiplicative model** the magnitude of the seasonal pattern depends on the magnitude/level of the data.

Decomposition: Seasonal Component

3. Assume the observed detrended series can be **represented** as

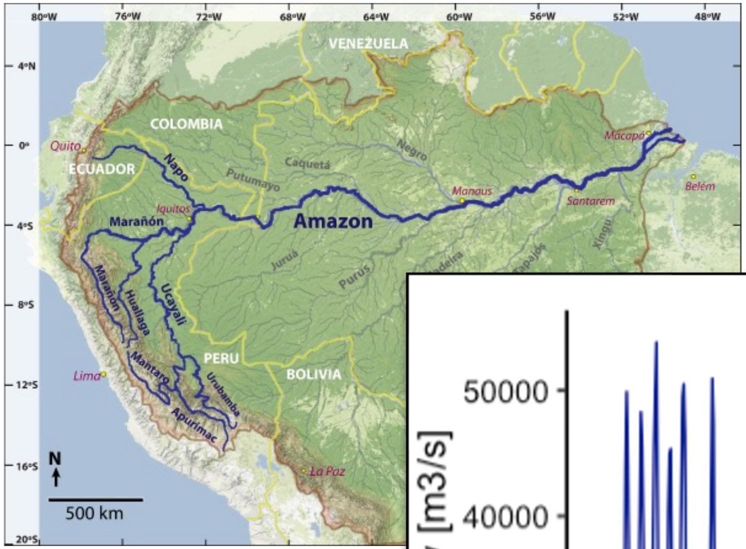
$$Y_{seasonal_t} = \sum_{s=1}^{12} \beta_s D_{s,t} + \epsilon_t \qquad \text{where } E[\epsilon_t] = 0$$

$$D_{s,t} = \begin{cases} 1 & if\ t\ belongs\ to\ season\ s \\ 0 & o.w. \end{cases} \qquad for\ s = 1,2, \dots 12$$

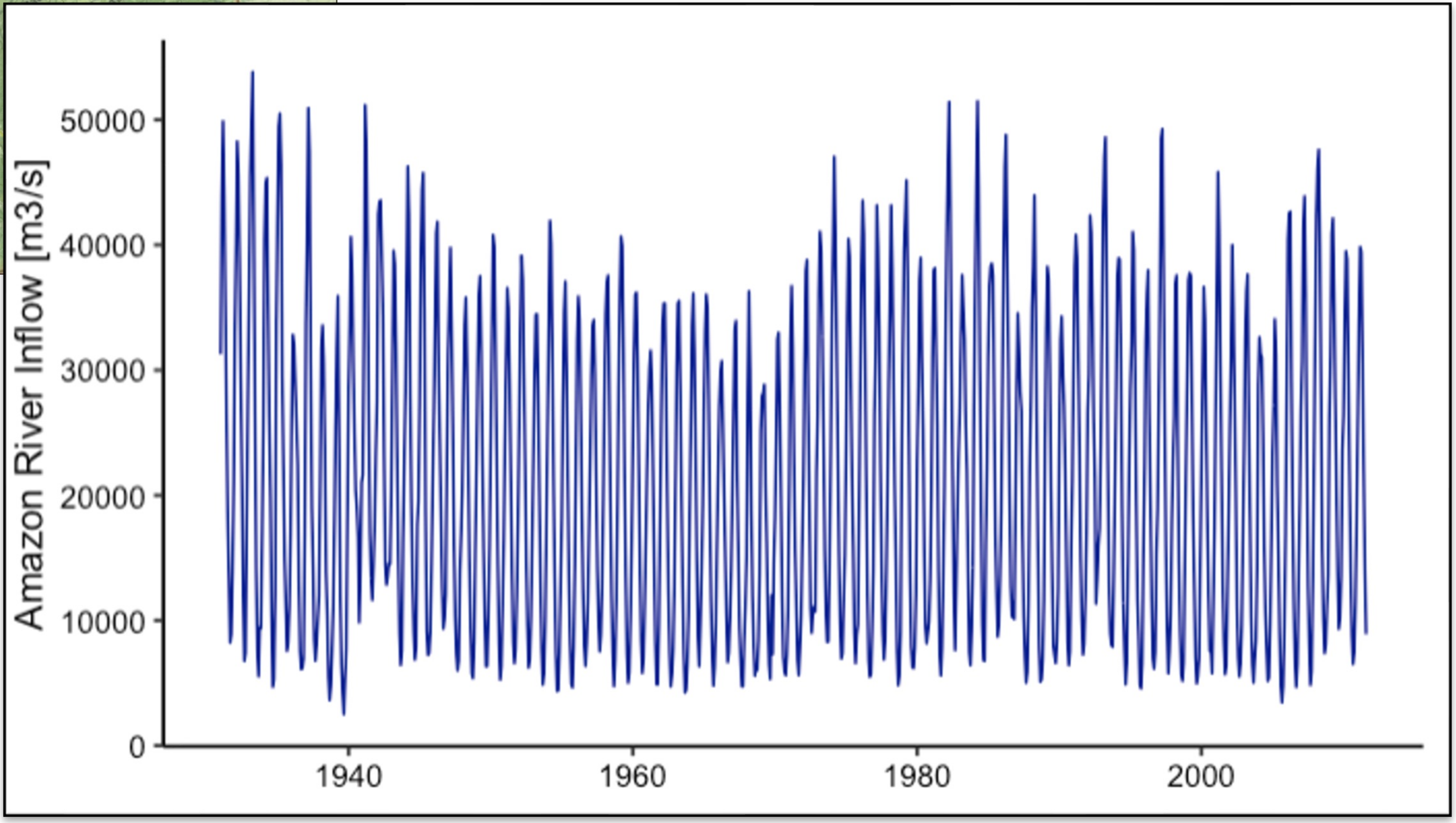
4. **Estimate** the parameters $\beta_1, \beta_2, \dots \beta_{12}$ by regression

Example: Inflow Data

Amazon River Inflow in m^3/s

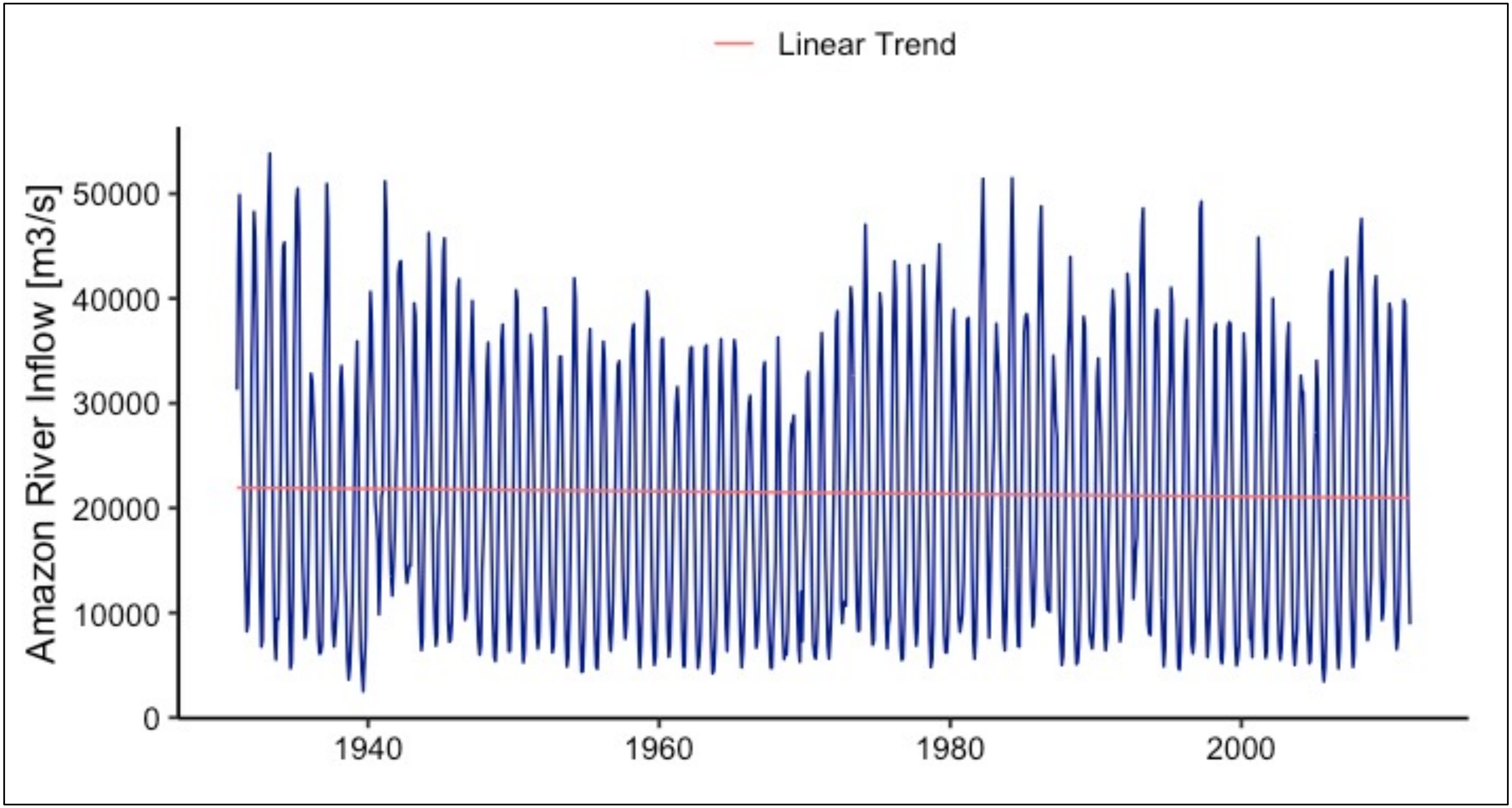


Which components/patterns can you see in this series?



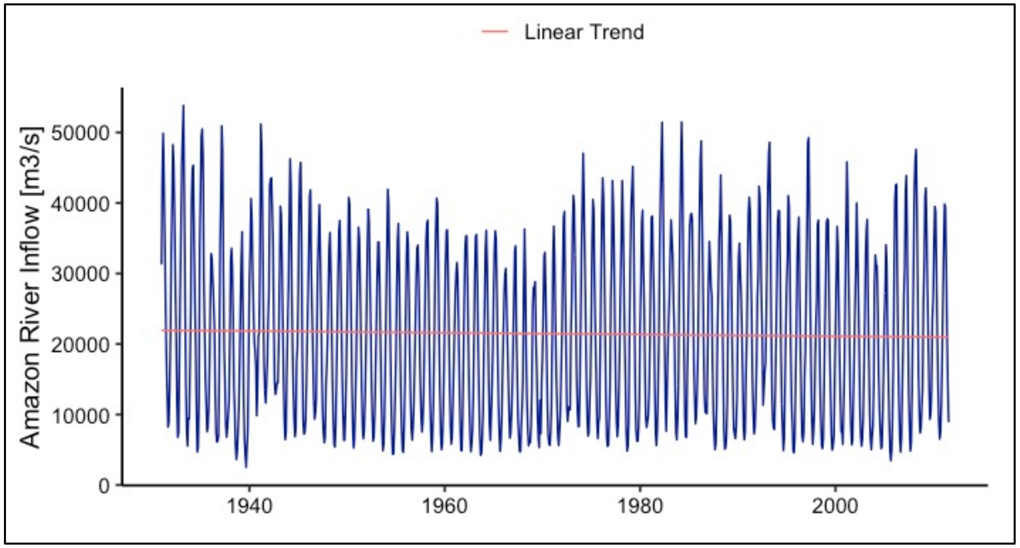
Linear Trend visualization

$$Y_{trend_t} = \beta_0 + \beta_1 t$$



Linear Trend regression results

$$Y_{trend_t} = \beta_0 + \beta_1 t$$

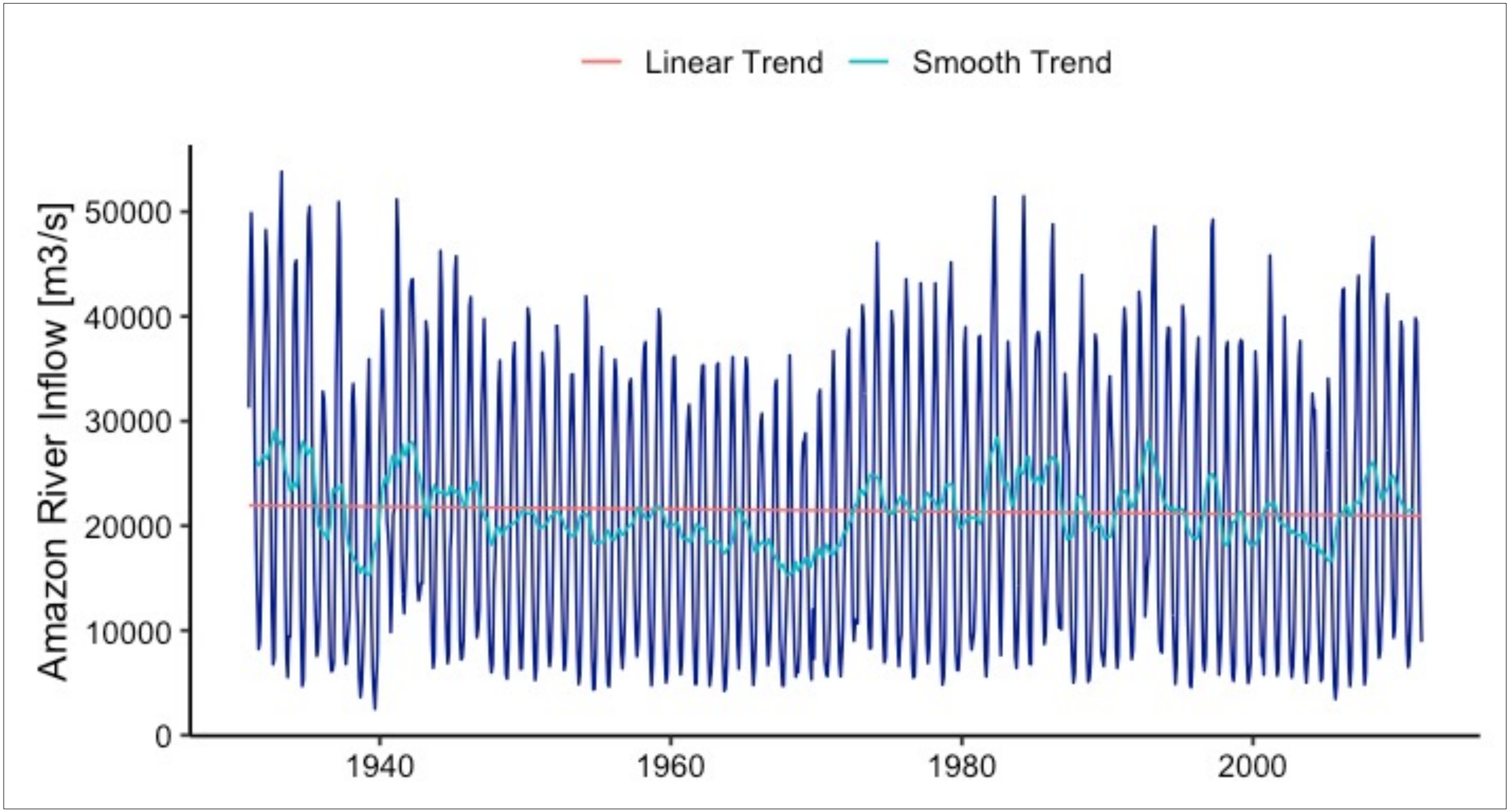



```
Call:
lm(formula = inflow ~ t, data = data)

Residuals:
   Min     1Q Median     3Q    Max
-19337 -11555  -1483  10061  31900

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 21949.403    797.962  27.507   <2e-16 ***
t              -1.024      1.427  -0.718    0.473
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 12400 on 966 degrees of freedom
Multiple R-squared:  0.0005328, Adjusted R-squared:  -0.0005018
F-statistic: 0.515 on 1 and 966 DF,  p-value: 0.4732
```

p-value > 0.05
Coefficient β_1 not significant

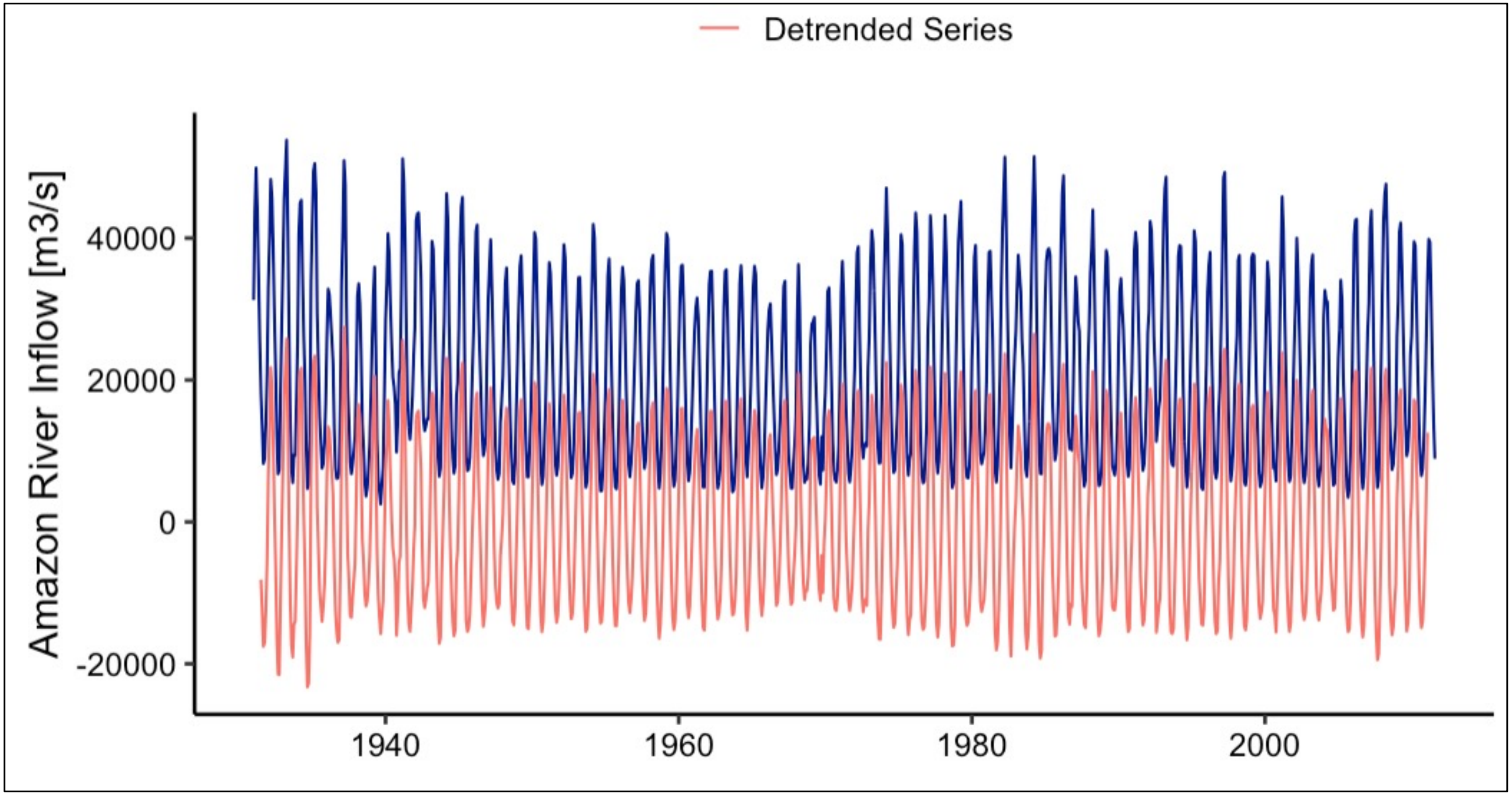
Smooth Trend: MA(6)

$$Y_{trend_t} = average(Y_{t-6} + Y_{t-5} + \cdots + Y_t + \cdots + \ Y_{t+5} + Y_{t+6})$$

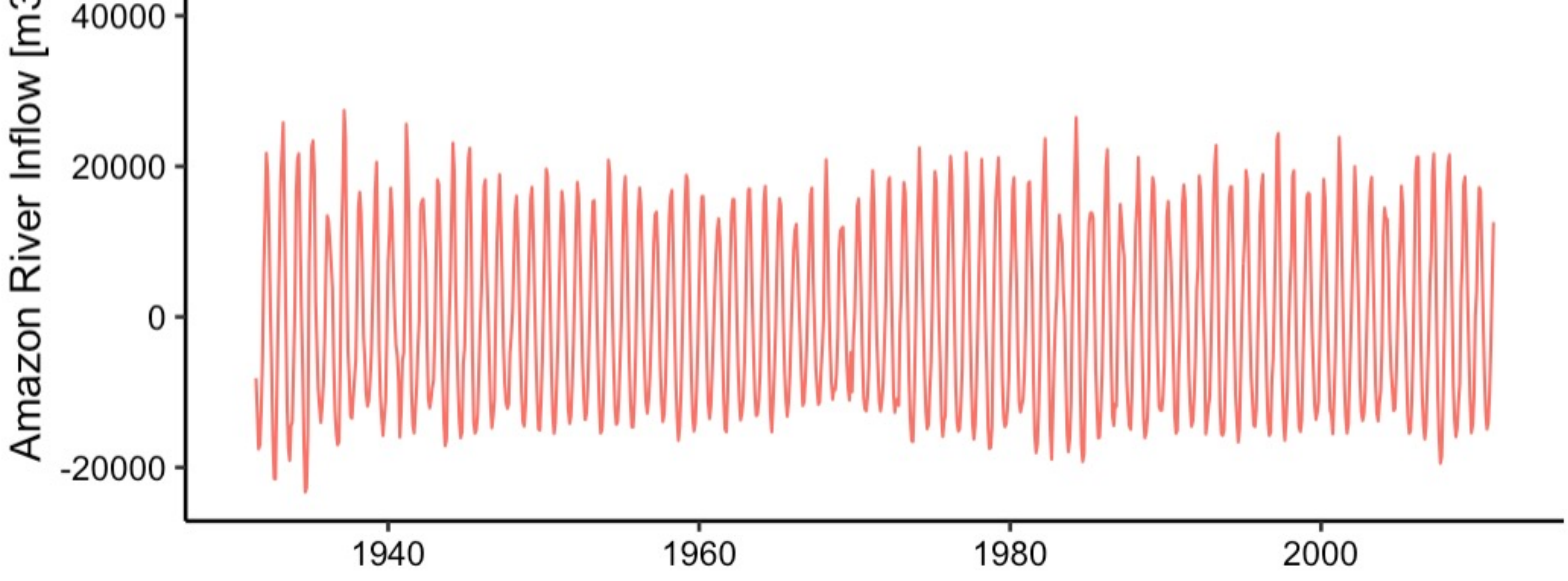


Detrended Series with Additive Model

$$Y_{detrend_t} = Y_t - Y_{trend_t}$$



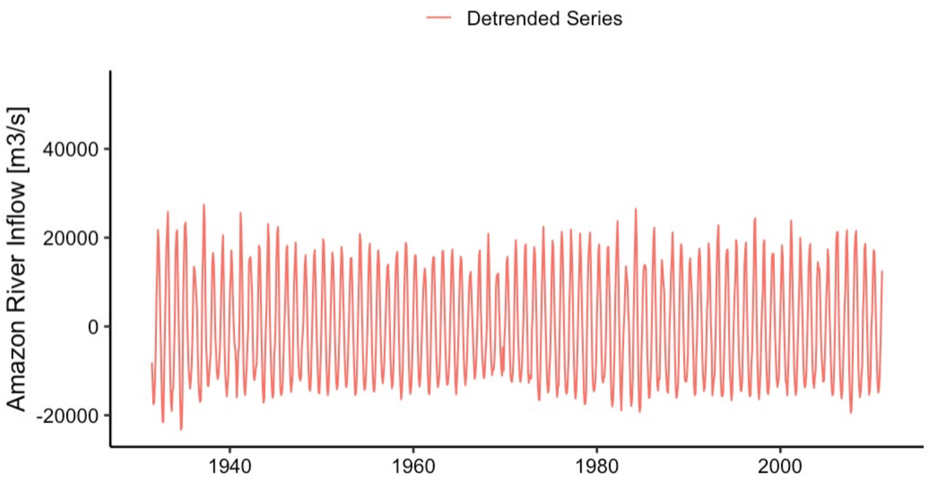
Do you still see any patterns?



Seasonal Component Regression Results

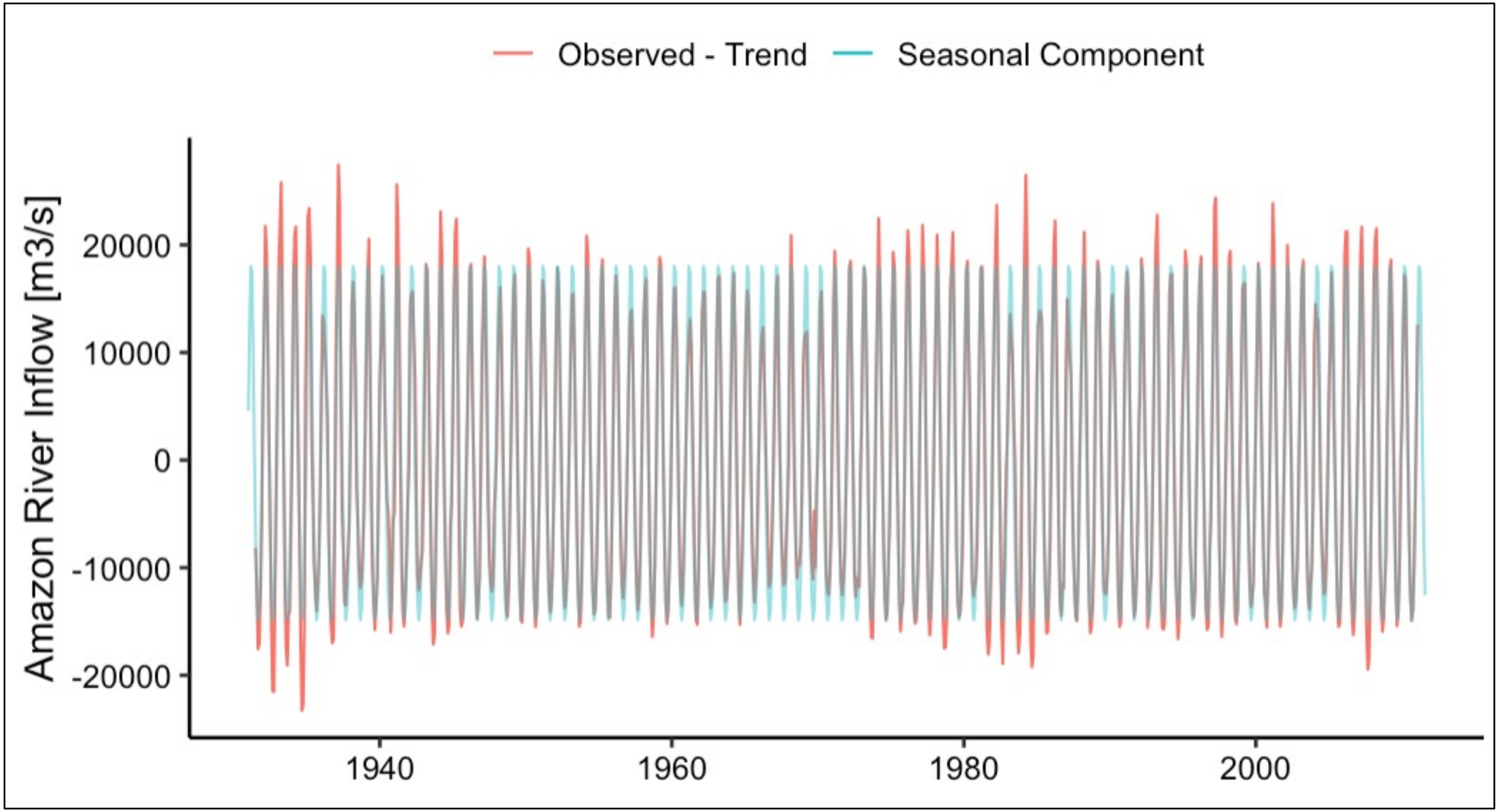
$$Y_{seasonal_t} = \sum_{s=1}^{12} \beta_s D_{s,t} + \epsilon_t$$

```
Call:
lm(formula = detrend_smooth ~ dummies)

Residuals:
   Min     1Q Median     3Q    Max 
 -9430  -1552    -50   1654  10797 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  -3545.7      307.4 -11.533   <2e-16 ***
dummiesJan    8131.5      434.8  18.702   <2e-16 ***
dummiesFeb   15948.6      434.8  36.682   <2e-16 ***
dummiesMar   21556.7      436.2  49.424   <2e-16 ***
dummiesApr   20984.3      436.2  48.112   <2e-16 ***
dummiesMay   12870.3      436.2  29.508   <2e-16 ***
dummiesJun    3959.1      436.2   9.077   <2e-16 ***
dummiesJul   -3731.7      434.8  -8.583   <2e-16 ***
dummiesAug   -8994.6      434.8 -20.688   <2e-16 ***
dummiesSep  -11306.9      434.8 -26.006   <2e-16 ***
dummiesOct  -10378.9      434.8 -23.872   <2e-16 ***
dummiesNov   -6515.4      434.8 -14.986   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2750 on 944 degrees of freedom
  (12 observations deleted due to missingness)
Multiple R-squared:  0.9478,    Adjusted R-squared:  0.9472 
F-statistic:  1558 on 11 and 944 DF,  p-value: < 2.2e-16
```

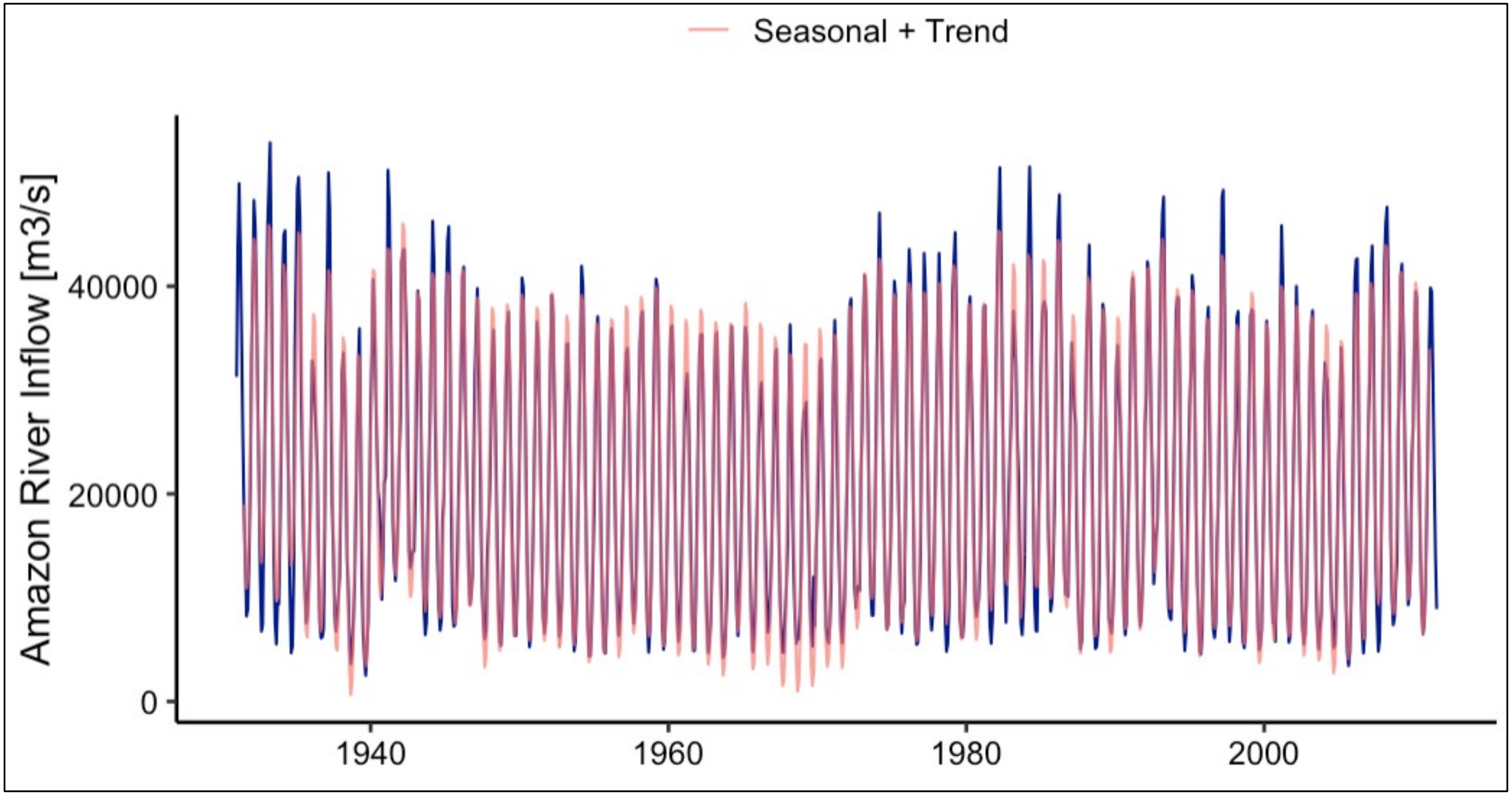


Seasonality explains most of the variability

Seasonal Component Visualization



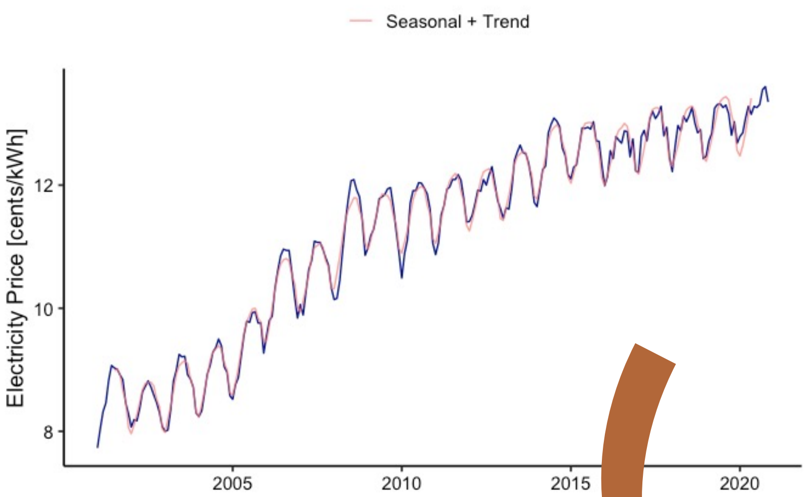
Seasonal + Trend Model of Inflow

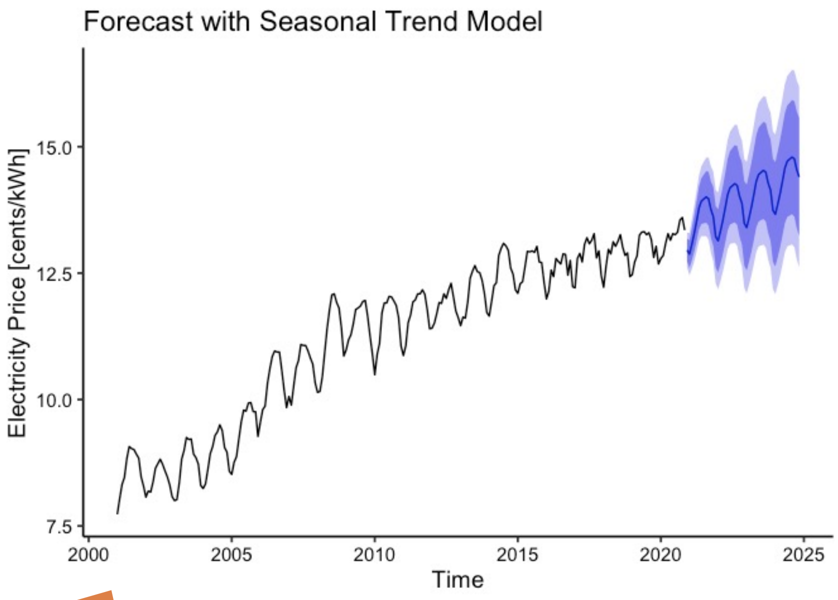


Recall dark blue line represents observed data

More Examples

- Electricity Price – STL model





Please visit: https://github.com/ENV790/TSA_Intro_EDAL

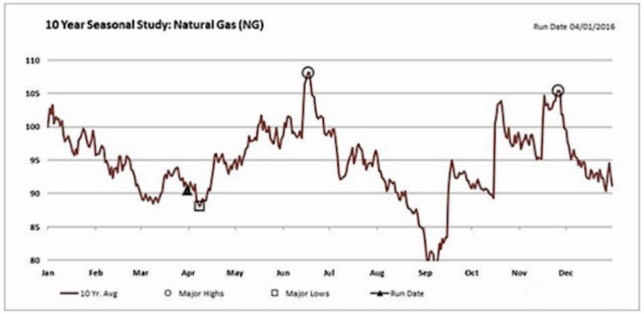
Forecasting and Model Selection

Time Series Forecasting

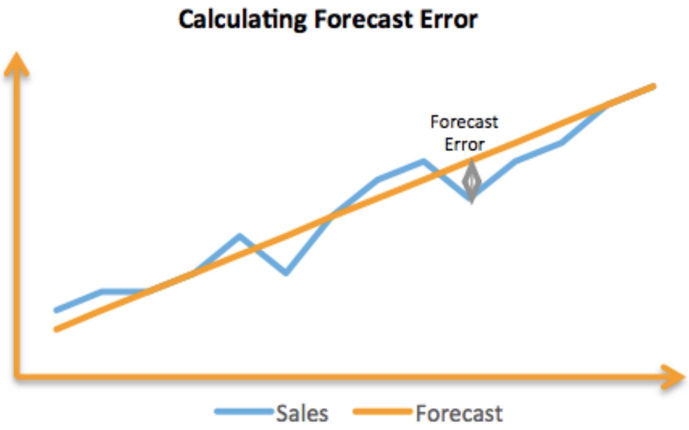
- **Forecast/Projection** → statement about the future value/probability distribution of a variable of interest
  - Forecasts are often used for weather, demand, and resource availability
  - Important element in making informed decisions
  - uses past values of a time series to estimate future values of the same time series

Two Important Aspects of Forecasts

- Expected **level**
  - Ex: the level of demand may be a function of some <u>structural variation</u> such as trend or seasonal variation



- **Accuracy**
  - Related to the potential size of forecast error



Features Common to All Forecasts

- Techniques assume some underlying **causal system that existed in the past will persist into the future**
- Forecasts are **not perfect**
  - Probability density function: provide an indication of the extent to which the forecast might deviate from the forecasted value
- Forecasts for **groups of items are more accurate** than those for individual items
- Forecast **accuracy decreases as the forecasting horizon increases**

Model Selection

"Unsolved" problem in statistics: there are no magic procedures to get you the "best model" (Kadane and Lazar)

- With a limited number of predictors, it is possible to search all possible models
- But when we have many predictors, it can be difficult to find "best" model (many possibilities)
- How do we select models?
  - We need a criteria or benchmark to compare two models
  - We need a search strategy

Model Selection

Which models best explain the observed data?

Which models give the best predictions of future observations?

- Backward-looking assessment
  - **Residual Analysis:** describes the difference between actual historical data and the fitted values generated by a statistical model
  - **Akaike Information Criterion (AIC):** balance goodness of fit and the simplicity of the model

- Forward-looking assessment
  - **Forecast accuracy** is an important forecasting technique selection criterion

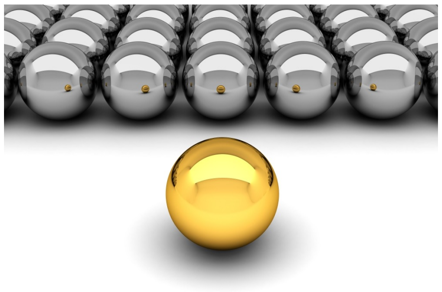
    Error = Actual – Forecast

    Observed value
  - **Root Mean Squared Error (RMSE)**
  - **Mean Absolute Percentage Error (MAPE)**
  - Need to hold-out some observed data to test forecast accuracy

Always combine metrics!

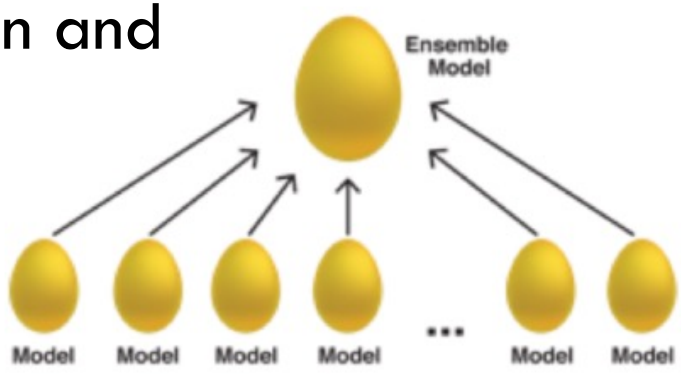
How to choose the best model?

- Build a set of candidate models and use **model selection/performance criteria** to pick one



- To improve forecasting accuracy, **combine forecasts** derived from models that differ in construction and predictor variables

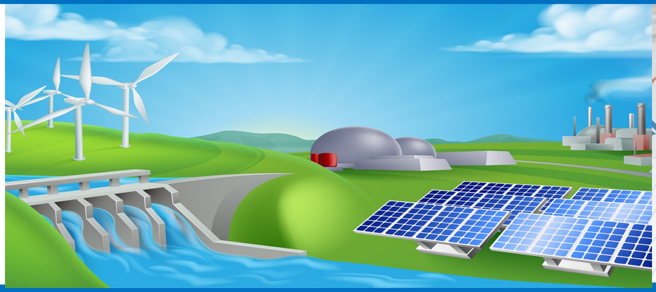
ENSEMBLE



Additional Resources

- Time Series Analysis for Energy and Environment Applications
  - Spring 2023 Course website: https://env790.github.io

- Environmental Data Analytics
  - Spring 2023 Course website: https://env872.github.io




THANK YOU !

luana.marangon.lima@duke.edu