Timestamp Data Analysis : ARIMA Modeling

Application on IPG2211A2N Dataset: Monthly Industrial Production Index for Electric and Gas Utilities

Albert Tchaptchet Womga

CIMT College
Supervised by: Priya Virdi

July 30, 2025

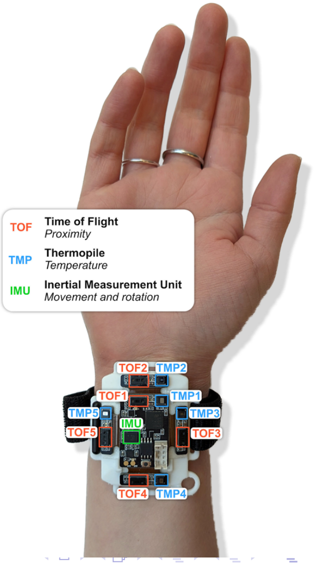
Outline

- **Timestamp:** A specific label indicating the exact time a data point was recorded. Formatted as `YYYY-MM-DD` or `YYYY-MM-DD HH:MM:SS`.
- **Role:** Enables ordering, aligning, and analyzing events over time.
- **Usage:** Critical in logs, sensors, finance, IoT.
- **Objective:** Enable time-indexed operations like filtering, resampling, lagging, forecasting.

Figure: Vintage T. Reiner timestamp machine, No. 650047, with "Date" labels

Helios Device: IoT Wearable Example

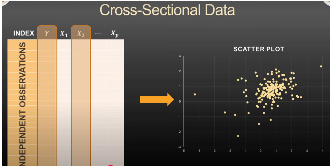
- CMI – Detect Behavior with Sensor Data
- Captures multi-modal time-series data for behavioral analysis.
- Sensors:
  - IMU (Accelerometer, Gyroscope, Magnetometer)
  - Thermopile (Temperature sensors)
  - ToF (Proximity sensors with 320 channels)



Time Series vs. Cross-Sectional

- **Time Series:** Ordered values over time.
- **Cross-Sectional:** Values captured at a single point in time. E.g., survey results.
- **Key Distinction:** Unlike cross-sectional data (snapshot at one point in time), time series values evolve over time.

Time Series Example



Cross-Sectional Example

Single vs. Multiple Time Series

- **Single Series:** One measurement over time.
- **Multiple Series:** Track several variables together (e.g., temperature + humidity).
- Useful in multivariate time series modeling.

Time Series Decomposition Concept

Key Idea: A time series can be decomposed into two main parts:

- **Signal** – The systematic component: Trend, Seasonality, Cycle.
- **Noise** – The irregular or random component.

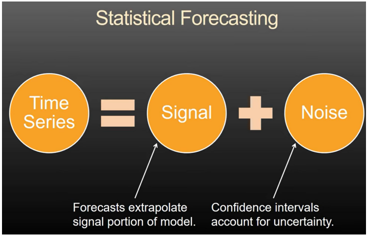


Figure: Decomposition = Signal + Noise

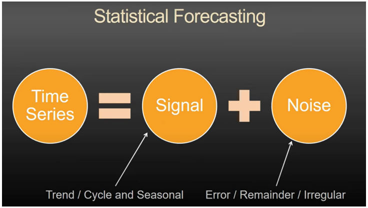
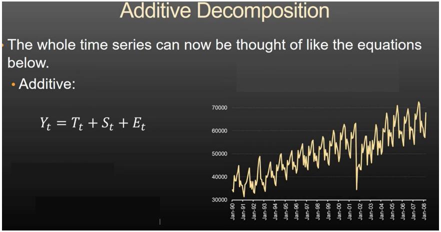
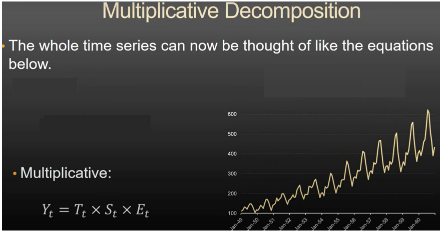


Figure: Signal: Trend + Seasonal — Noise: Irregularity

Additive vs Multiplicative Decomposition





- **Additive:** Series $=$ Trend $+$ Seasonality $+$ Noise
- **Multiplicative:** Series $=$ Trend $\times$ Seasonality $\times$ Noise
- Use additive when seasonal fluctuations are constant; multiplicative when they grow with trend.

ARIMA: Core Components

ARIMA combines three techniques for time series forecasting:

ARIMA(p,d,q) Notation

- **p**: Autoregressive terms (memory of past values)
- **d**: Differencing degree (makes series stationary)
- **q**: Moving average terms (memory of past errors)

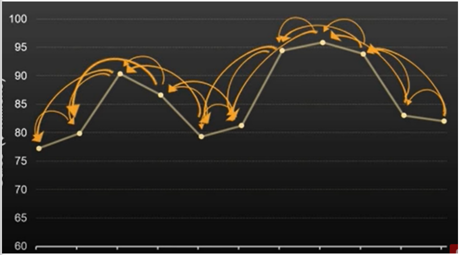
$$\text{ARIMA} = \underbrace{\text{AR}}_{\text{p}} + \underbrace{\text{I}}_{\text{d}} + \underbrace{\text{MA}}_{\text{q}}$$

Autoregressive (AR) Component

AR(p) Model

$$y_t = c + \sum_{i=1}^{p} \phi_i y_{t-i} + \epsilon_t$$

- Uses *past values* to predict future.
- ϕ_i: Autoregressive coefficients.
- PACF (Partial Autocorrelation Function) helps identify order p.



Integrated (I) Component

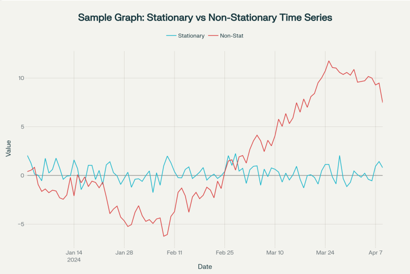
Differencing makes series stationary:

First Difference:

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

Second Difference:

$$\Delta^2 y_t = \Delta y_t - \Delta y_{t-1}$$



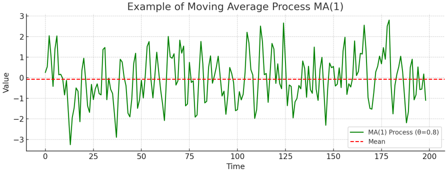
- d = Number of differencing operations
- Removes trends/seasonality
- Stationarity confirmed by ADF test

Moving Average (MA) Component

MA(q) model:

$$y_t = c + \epsilon_t + \sum_{j=1}^{q} \theta_j \epsilon_{t-j}$$

- Uses *past forecast errors*
- θ: MA coefficients
- ACF helps identify q
- **MA(1)** is stationary by definition



ARIMA Model Equation

General **ARIMA(p,d,q)** formulation:

$$(1 - \sum_{i=1}^{p} \phi_i B^i)(1 - B)^d y_t = c + (1 + \sum_{j=1}^{q} \theta_j B^j)\epsilon_t$$

ARIMA(1,1,1) Example:

$$\underbrace{(1 - \phi_1 B)}_{\text{AR(1)}} \underbrace{(1 - B)}_{\text{I(1)}} y_t = c + \underbrace{(1 + \theta_1 B)}_{\text{MA(1)}} \epsilon_t$$

Which expands to:

$$\Delta y_t = c + \phi_1 \Delta y_{t-1} + \theta_1 \epsilon_{t-1} + \epsilon_t$$

- B: Backshift operator ($B^k y_t = y_{t-k}$)

Choosing p,d,q:

- d: Minimum differencing for stationarity (ADF test)
- p: PACF cutoff lag
- q: ACF cutoff lag
- Use AIC/BIC for model comparison

When to use ARIMA:

- Non-seasonal data with trends
- Univariate time series
- Short-term forecasting

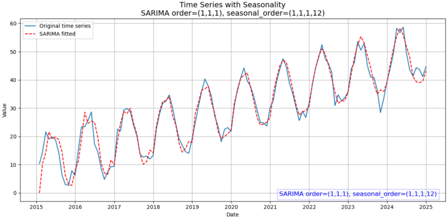
Limitations:

- Not for volatility (use GARCH)
- Not for multivariate series (use VAR)

ARIMA Extensions

Common Variations

- **SARIMA**: Seasonal ARIMA (e.g., ARIMA$(p, d, q)(P, D, Q)_s$)
- **ARIMAX**: With exogenous variables
- **RegARIMA**: With regression components



ARIMA + Seasonality = SARIMA

Dataset Overview

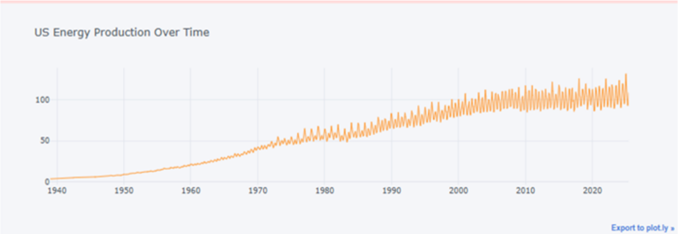
About the Dataset

- **Source:** FRED - U.S. Federal Reserve Economic Data.
- **Link:** `https://fred.stlouisfed.org/series/IPG2211A2N`
- **Variable:** Monthly Industrial Production Index for electric and gas utilities.
- **Coverage:** U.S.-based establishments from 1939 to present.

Data Preprocessing

Steps Performed

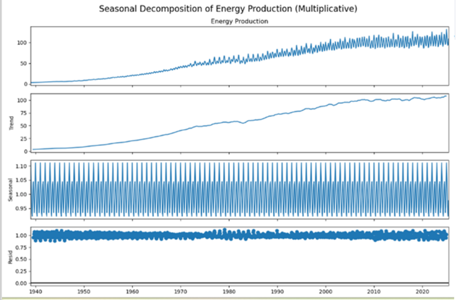
- Converted `observation_date` to datetime.
- Set datetime as index.
- Renamed value column to `Energy Production`.



Seasonal Decomposition: Multiplicative Model

Decomposition Components

- **Observed:** Raw data showing annual oscillations and long-term growth.
- **Trend:** Strong rise until early 2000s, then flattens.
- **Seasonal:** Stable yearly cycle tied to heating/cooling demand.
- **Residual:** Minimal, indicating good model fit.



Interpretation of Decomposition

Detailed Analysis

- **Observed:** Upward trend (1940s–2000), oscillations, mild decline post-2008.
- **Trend:** Long-term increase followed by stabilization.
- **Seasonal:** Repeats every 12 months, stable across decades.
- **Residual:** Random noise near 1.0, no major outliers.

Conclusion

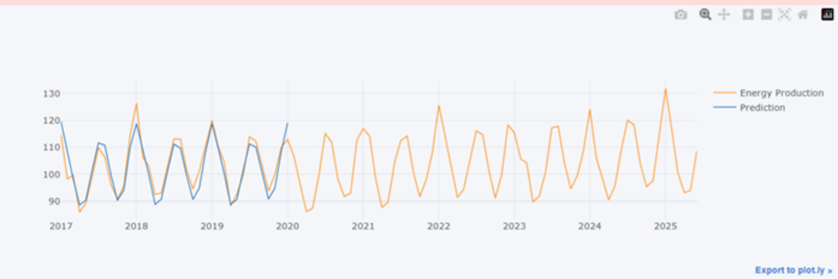
Strong trend and seasonality captured by multiplicative model. Residuals confirm good fit.

Seasonal ARIMA Modeling

Modeling Approach

- Used `pmdarima`'s `auto_arima` for parameter selection.
- Explored combinations of p, d, q, P, D, Q.
- Optimized for minimum AIC.
- **Best Model:** $ARIMA(1,1,1)(2,1,2)[12]$
- **Best Model AIC:** 4228.12

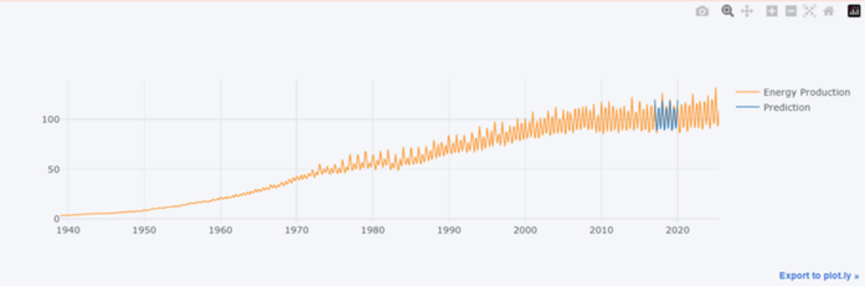
Forecast Plot: Focused 2017–2025



Insight

Forecast closely tracks seasonal trends and aligns with actual values from 2017–2019.

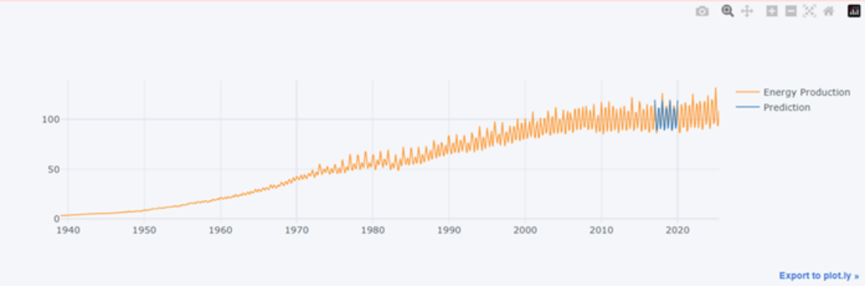
Forecast Plot: Full Timeline (1940–2025)



Insight

Model captures long-term growth and cycles accurately.

Forecast Plot: Full Timeline (1940–2025)



Insight

Model captures long-term growth and cycles accurately.

Observed vs Forecasted Energy Production

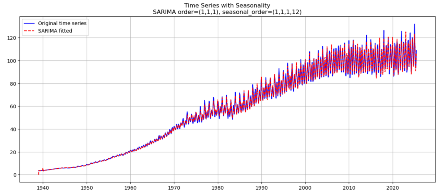


Figure: Observed Energy Production with SARIMA Fit (1940–2025)

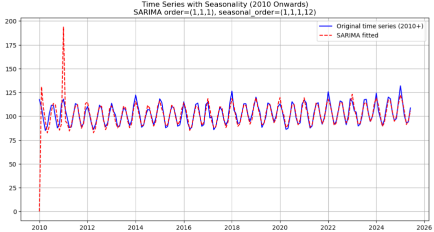


Figure: Zoomed Forecast (2010–2025) with SARIMA Fit

From the Plots

The left plot shows the full historical dataset with SARIMA fitted values closely tracking the observed data. The right plot zooms in from 2010 onwards, where the model demonstrates excellent short-term tracking and reliable forecasting for future energy production.

Forecast Details

- **Model:** SARIMA(1,1,1)(1,1,1,12)
- **Data:** Period from 2010 onwards.
- The model captures both trend and strong seasonal cycles.
- Forecast extends 36 months ahead with 95% confidence intervals.
- Observed data (blue), forecast values (green), and uncertainty bands (light green).

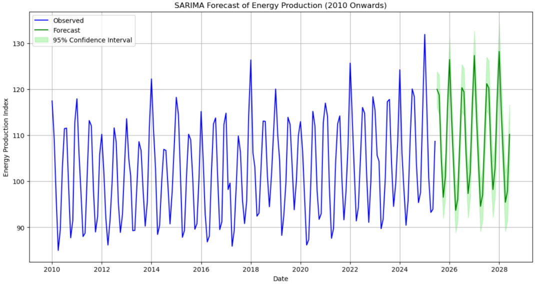


Figure: SARIMA forecast with confidence intervals.

Conclusion

- The energy production index exhibits strong seasonality and long-term structural growth.
- Multiplicative decomposition was appropriate and showed excellent model fit.
- The ARIMA model effectively captured both trend and seasonality for accurate forecasting.

Future Scope

- Integrate exogenous variables (e.g., temperature, policy changes) using ARIMAX.
- Apply deep learning models (e.g., LSTM, GRU, RNN , CNN models) for non-linear dependencies.
- Expand to real-time forecasting and anomaly detection in smart grids.

References

- Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and Practice*.
- Box, G. E. P., Jenkins, G. M., & Reinsel, G. C. (2015). *Time Series Analysis: Forecasting and Control*.
- FRED – Federal Reserve Economic Data: `https://fred.stlouisfed.org/series/IPG2211A2N`
- Statsmodels documentation: `https://www.statsmodels.org/`
- Pmdarima documentation: `https://alkaline-ml.com/pmdarima/`
- Child Mind Institute. (2024). Helios: A Wearable Device for Behavioral Health Monitoring. `https://www.childmind.org/helios`