Bachelor's Project
Power Grid Load Forecasting using Machine Learning Approaches

Omar Abdesslem
University of Geneva

April 2025

Contents

1 Abstract

This project focuses on developing machine learning models for load forecasting in the Swiss energy grid, using historical data on energy consumption, production, and cross-border exchanges. The datasets include detailed information on total energy consumed and produced in the Swiss control block, grid feed-ins, net outflows, and energy trades with neighboring countries (Germany, France, Austria, and Italy). By leveraging this data along with weather and seasonal factors, the project aims to improve the accuracy of short-term and long-term load forecasts using advanced machine learning techniques such as LSTM, Transformers, and Gradient Boosting, while comparing their performance with traditional models.

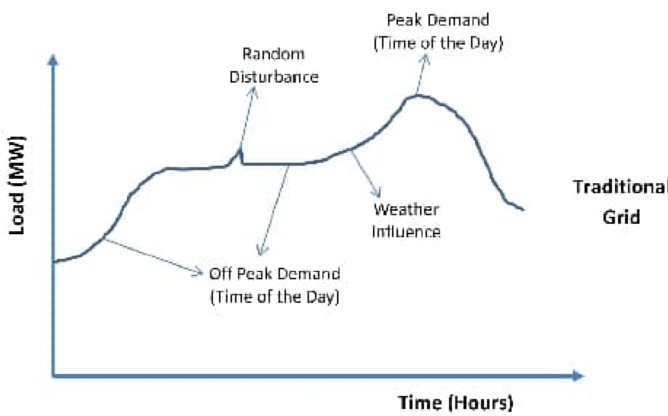


Figure 1: Load and influence

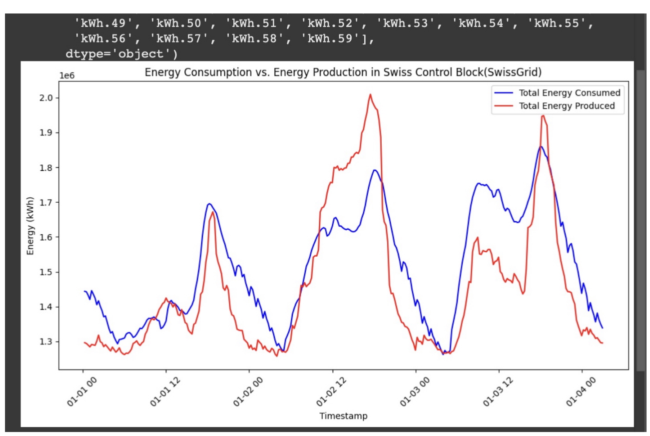


Figure 2: Plotted Energy Production visuals (Python3)

The goal of which is to combine Machine Learning, Data Structures, and Physics to predict real-life energy trends.

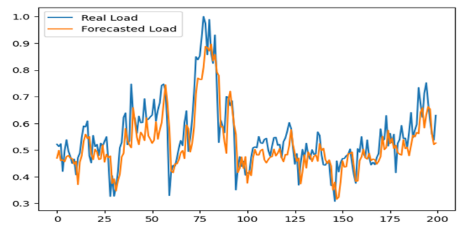


Figure 3: Endgoal

2 Objectives

The primary objectives are:

1. Visualize the Data

2. Develop a Medium Term Forecasting Model (Predicting Total Amount of Energy Consumed per day/week)

3. Setting a Baseline Model and model evaluation Metric

4. Evaluate the results

5. Discussing challenges and conclusions

3 Visualisation

3.1 Variable to Predict

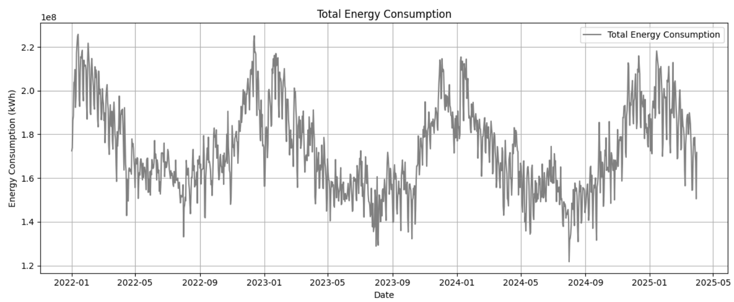


Figure 4: Total Swiss Energy Consumption(2024)

4 Time Series Analysis

Data Cleaning

will be added in: for now: converted everything to numeric, replaced errors with NaN, removed NaN rows

Time Series Analysis

4.1 Mathematical basics

4.1.1 Stationarity

Stationarity refers to the behavioral consistency of the time series. Mathematically, this means that the mean and covariance stay invariant regardless of the time shift.

Strict stationarity means that the mean, variance, and covariance are constant. Weak stationarity means that the mean, variance is constant, and the covariance function $\gamma(s, t)$ depends only on t - s, as in any two values depends only on the time difference between them, not on the actual time at which they occur. [4]

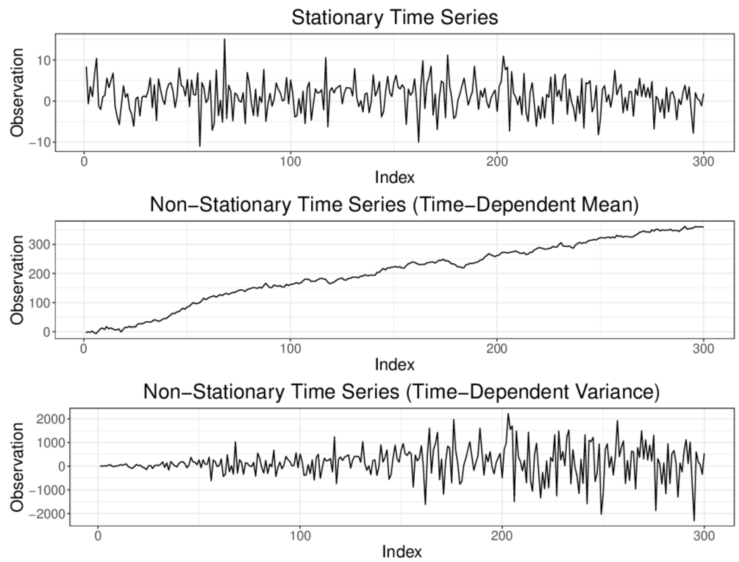


Figure 5: Stationary and non-stationary time series, Bauer 2021 [1]

4.1.2 White noise

A stochastic process $\{Y_t\}$ is called *white noise* if all its elements are uncorrelated, with mean $E(Y_t) = 0$ and variance $\text{Var}(Y_t) = \sigma^2$ [4]. The standard deviation is measured by [16]:

$$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2}$$

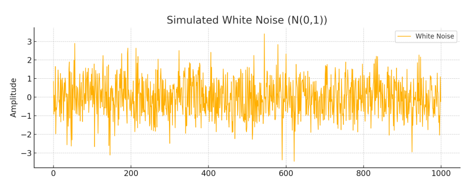


Figure 6: White noise using an np.random fct

As explained in Wikipedia [16], white noise is a stochastic process and does not have a deterministic function.

4.1.3 Random Walk

A time series $\{Y_t\}$ is called a *random walk* if it satisfies the relation

$$Y_t = Y_{t-1} + \varepsilon_t,$$

where ε_t is white noise. [4]

4.2 Initial data analysis

4.2.1 Mean

The mean, or expected value, of a time series quantifies its average level over time. Its formula is given by:

$$\mu = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Where x_i are the observed values and n is the number of observations [2].

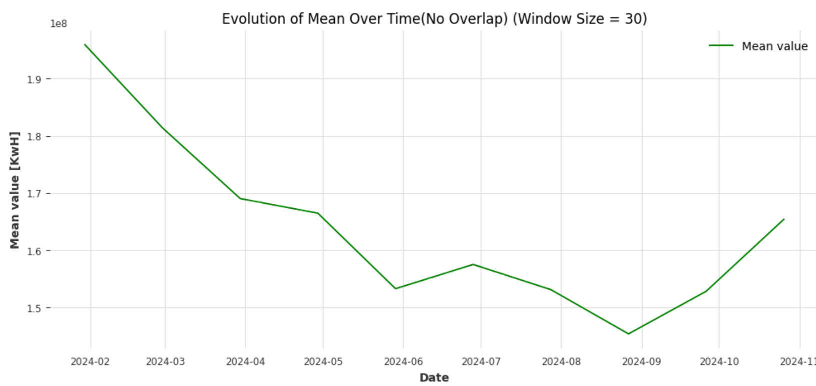


Figure 7: Evolution of Mean Over Time — 2024

In this case, the mean energy consumption shows a decreasing trend from February to September 2024, followed by a gradual increase into October. This might be due to seasonal changes (for example, warmer months might require less energy consumption than colder ones).

4.2.2 Variance

The variance measures the dispersion or spread of the data around the mean. Its formula is given by:

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2$$

Where μ is the mean and σ^2 is the variance [3].

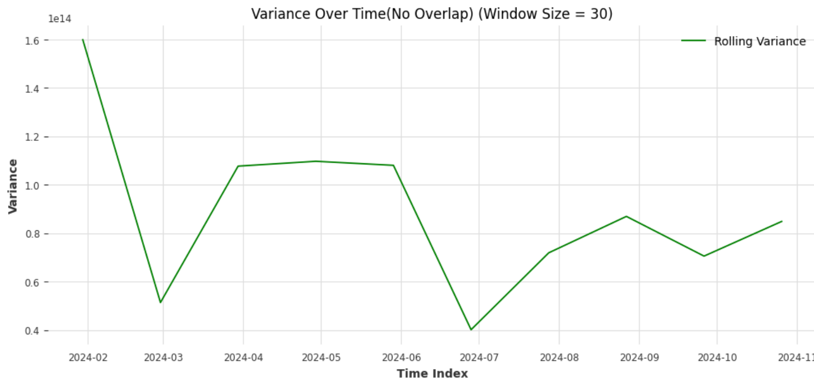


Figure 8: Evolution of Variance Over Time — Total Energy Consumption 2024

Observation: The variance fluctuates significantly over time.

4.2.3 Transformation

4.2.4 Moving Sum (Window)

The moving sumcalculate the sum of a fixed number of consecutive values (a "window") in a dataset. The Moving Sum formula is calculated by:

$$\mathrm{S}_t = \sum_{i=0}^{n-1} x_{t-i}$$

In this case, To reduce noise and reveal trends, we will be using a 1 week window.

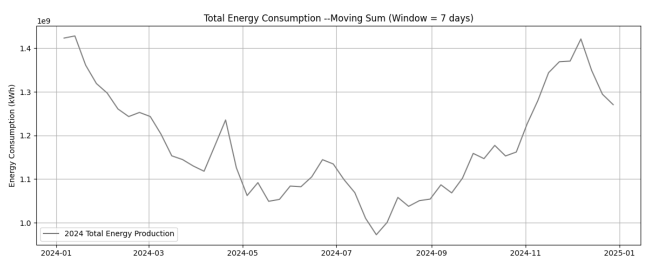


Figure 9: Weekly Sum

4.2.5 Moving Average

Averaging reduces variance, and introduces correlation in Yt [4]. A Simple Moving Average is calculated by the formula:

$$\mathrm{MA}_t = \frac{1}{n} \sum_{i=0}^{n-1} x_{t-i}$$

4.2.6 Outliers & Data Cleaning

An outlier is an observation that causes surprise relative to the rest of the data. It may be isolated or successive [4].

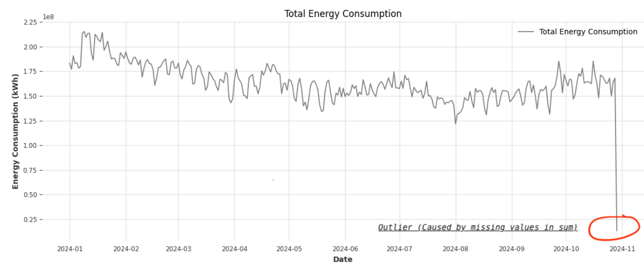


Figure 10: Outlier example in 2024 Consumption Graph

In the case for outliers, I will replace them with the value of the average in the Moving Average Window.

4.2.7 Trends

Trend is a pattern in data that shows the movement of a series to relatively higher or lower values over a long period of time [9]. Trends can be linear, quadratic, periodic, or more complex [4].

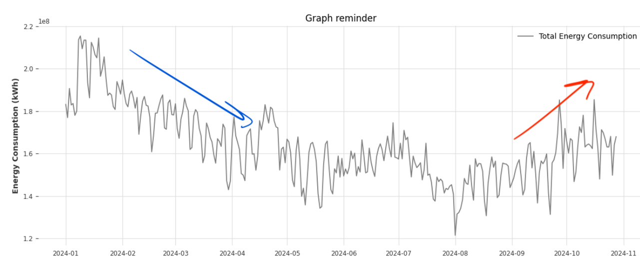


Figure 11: Trend Evolution in 2024 Consumption Graph

4.2.8 Seasonality

A repeating pattern that occurs at fixed and regular intervals (e.g. daily, weekly, yearly) [4]. Seasonality is a predictable cyclical pattern, whereas trends are a long-term change in data (increase/decrease).

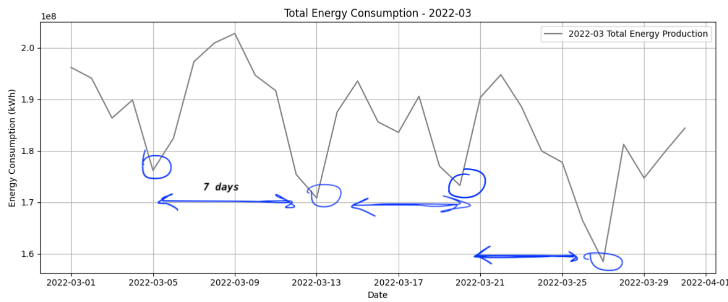


Figure 12: Weekly Seasonality, Month of March 2022

4.3 Differencing

4.3.1 Motivation

Differencing is a simple approach to removing trends. No need to estimate parameters. [4].

4.3.2 Differencing types

Differecing can be of *first-order* or *higher-order* [4]

4.3.3 First-order difference

The first order difference is defined as [4]:

$$\Delta Y_t = Y_t - Y_{t-1}$$

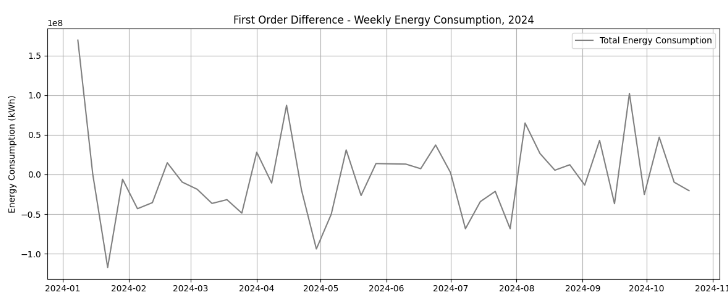


Figure 13: First Order Difference Weekly, 2024

Observation: The resulting time series look a lot more stable, less trends and maybe a mean revolving around zero? We will plot the new mean over time function to establish if it has become more stable.

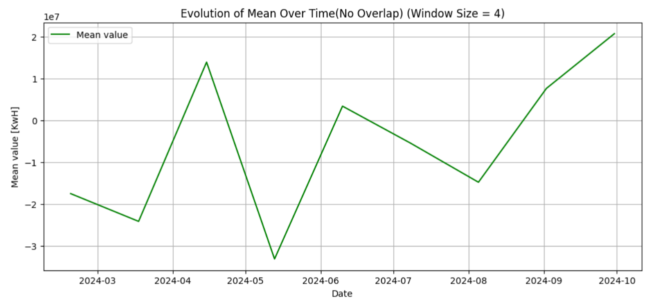


Figure 14: Mean Over Time after Differencing

4.3.4 Higher-order difference

If one round of differencing is not sufficient to achieve stationarity, a *higher-order difference* can be applied, the second-order difference is [4]:

$$\begin{aligned}\Delta^2 Y_t &= \Delta(\Delta Y_t) = \Delta(Y_t - Y_{t-1}) \\ &= \Delta Y_t - \Delta Y_{t-1} = Y_t - 2Y_{t-1} + Y_{t-2}\end{aligned}$$

First-order differencing reduces a random walk to stationarity. In practice, we difference until plots of the differenced data appear stationary; often k= 1,2 suffices [4].

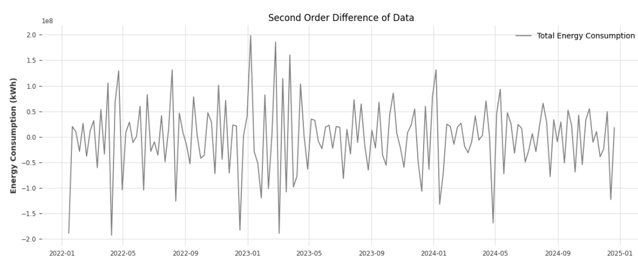


Figure 15: Second Order Difference Weekly, 2024

4.3.5 Seasonal differencing

4.4 ACF and PACF

4.4.1 Motivation

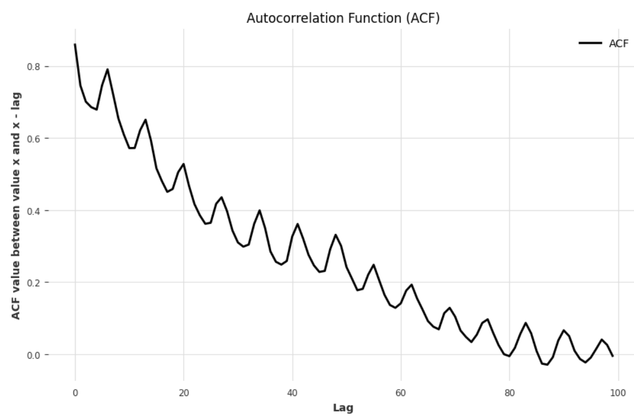
Autocorrelation and partial autocorrelation functions are used to understand the dependence structure of a time series. They help identify appropriate models [4].

4.4.2 Correlogram

The covariance function for equally spaced data $y_1, \ldots, y_n$ is defined as:

$$c_h = \frac{1}{n-h-1} \sum_{i=1}^{n-h} (y_i - \bar{y})(y_{i+h} - \bar{y}), \quad h = 0, 1, \ldots, n-2,$$

where $\bar{y}$ is the sample mean. The correlogram (ACF) is a graph of $\hat{\rho}_h = \frac{c_h}{c_0}$ against lag h [4].



1

Figure 16: AutoCorrelation Function 2024 - Total Energy Consumption

4.4.3 Partial correlogram

The PACF is interpreted in a similar way to the ACF, but it reveals the **direct** relationship between an observation and its lagged values, controlling for the values in between. Let $Y_0, \ldots, Y_h$ be successive observations. The partial autocorrelation function (PACF) at lag h measures the correlation between Y_t and Y_{t-h} after removing the linear influence of intermediate lags $Y_{t-1}, \ldots, Y_{t-h+1}$.

$$\tilde{\rho}_1 = \mathrm{corr}(Y_1, Y_0), \quad .$$

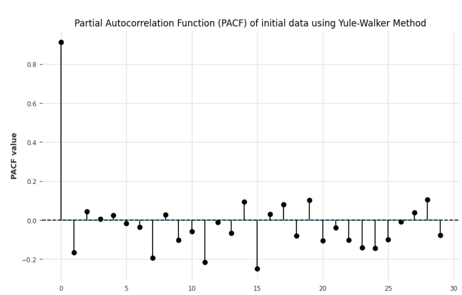


Figure 17: PACF computed using the Yule-Walker method

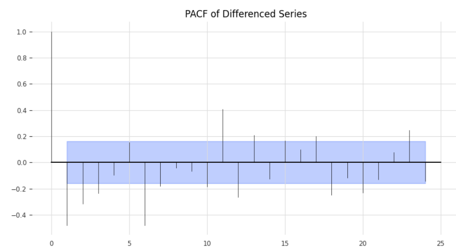


Figure 18: PACF generated using Darts

4.4.4 Testing for stationarity

To build reliable model, we need to check whether the data is stationary. One way to test this is to decompose the time series Y_t into three components:

$$Y_t = \xi_t + \eta_t + \varepsilon_t$$

[4]

Where:

- ξ_t is the deterministic trend, as in a fixed, predictable pattern over time
- η_t is a supposed random walk,
- ε_t is noise

Types of stationarity:

- **Level stationarity**: If $\sigma_u^2 = 0$ and $\xi_t = 0$, then Y_t is stationary around a constant mean.
- **Trend stationarity**: If $\sigma_u^2 = 0$ and $\xi_t = \beta t$, then Y_t becomes stationary after removing the trend.

KPSS Test: The KPSS test is used to test the null hypothesis that a time series is stationary. It does this by estimating the test statistic:

$$C(l) = \frac{1}{\sigma^2(l)} \sum_{t=1}^{n} S_t^2, \quad \text{where } S_t = \sum_{j=1}^{t} e_j$$

Here, $e_1, \ldots, e_n$ are the residuals from regressing Y_t on a constant or a linear trend (depending on whether testing for level or trend stationarity), and $\sigma^2(l)$ is a long-run variance estimate using a truncation lag l. [4] The test is interpreted as follows:

- If the test statistic is small (below the critical value), we **do not reject** the null hypothesis: the series is stationary.

- If the test statistic is large (above the critical value), we **reject** the null hypothesis: the series likely contains a unit root and is non-stationary.

```
is_stationary = stationarity_test_kpss(differed_series)
stat, p_value, lags, crit_vals = stationarity_test_kpss(differed_series)
print(f"KPSS statistic: {stat}"), print(f"p-value: {p_value}"), print(f"Is

✓ 0.0s                                                          Python
KPSS statistic: 0.19653700780101377
p-value: 0.1
Is stationary: True
```

Figure 19: KPSS result for differenced data

4.4.5 Testing for white noise

There are many methods to test for white noise. One of which, documented in the Time Series Analysis book [4], is the Ljung–Box.

For a time series $y_1, \ldots, y_n$, the Ljung–Box test statistic is:

$$Q_m = n(n+2) \sum_{h=1}^{m} \frac{\hat{\rho}_h^2}{n-h}$$

Where: - $\hat{\rho}_h$ is the autocorrelation of the sample at lag h - n is the length of the series - m is the maximum lag to include in the test [4]

We shall use the calculated autocovariances to create the Ljung-Box function using the formula above. The ACFs were calculated using biased covariances.

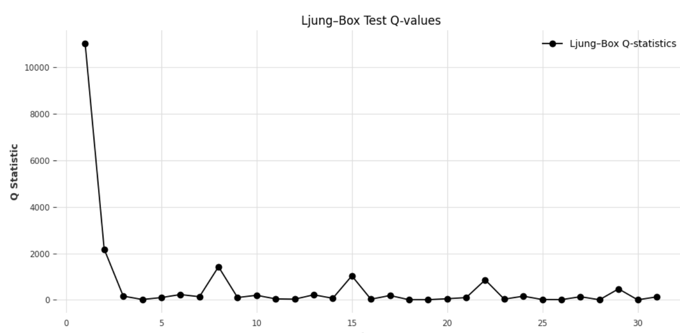


Figure 20: Enter Caption

Chi-squared distribution. Under the null hypothesis that the series is white noise, the Ljung–Box statistic Q_m follows a Chi-squared distribution with m degrees of freedom:

$$Q_m \sim \chi^2_m$$

The Chi-squared distribution is a continuous probability distribution. Its formula is the following for the squares of k independent standard normal variables:

$$\chi_k^2 = Z_1^2 + Z_2^2 + \cdots + Z_k^2, \quad \text{where } Z_i \sim \mathcal{N}(0,1)$$

Its probability density function is given by:

$$f(x;k) = \frac{1}{2^{k/2}\Gamma(k/2)} x^{(k/2)-1} e^{-x/2}, \quad x > 0$$

where Γ is the gamma function, as in the non-integer and integer factorial calculator. The distribution is positively skewed, and its shape depends on the degrees of freedom k.

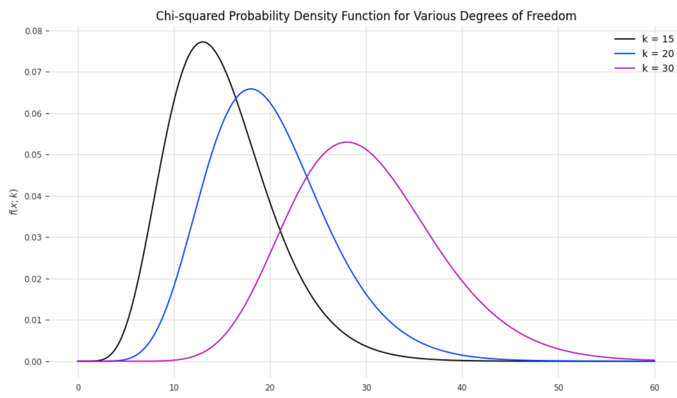


Figure 21: Enter Caption

In the Ljung–Box test, if the observed Q_m is greater than the critical value from the χ_m^2 distribution at a given significance level (e.g., 5%), we reject the null hypothesis and conclude that the time series is not white noise.

Then, the p-value is:

$$\text{p-value} = P(\chi_m^2 \geq Q_m) = 1 - F(Q_m)$$

Where:

$F(Q_m)$ is the distribution function of the Chi-squared distribution.

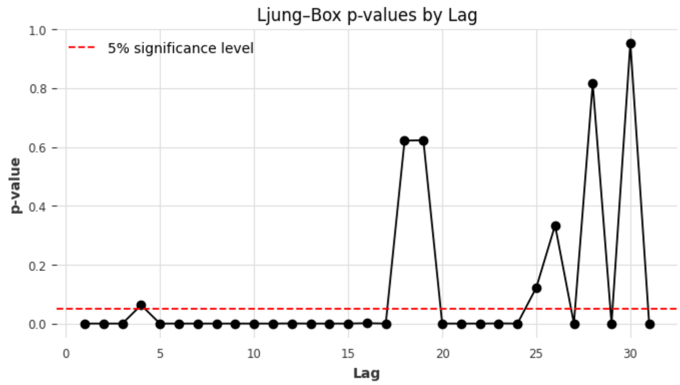


Figure 22: Enter Caption

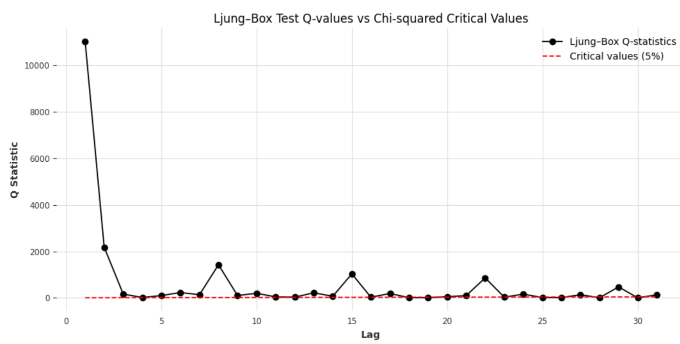


Figure 23: Enter Caption

4.4.6 Is this brownian noise

4.4.7 Checking normality using QQ plots

We often need to compare data $y_1, \ldots, y_n$ with a given distribution F, usually the normal distribution (for example, to check if the standardized residuals are $\mathcal{N}(0, 1)$).

A quantile-quantile (Q–Q) plot is a graph of the ordered values of the y_j:

$$y_{(1)} \leq y_{(2)} \leq \cdots \leq y_{(n)}$$

against theoretical quantiles of F, given by $x_i = F^{-1}\left(\frac{i}{n+1}\right)$: we plot pairs

$$(x_1, y_{(1)}), (x_2, y_{(2)}), \ldots, (x_n, y_{(n)})$$

It is best if the plot is square and if it includes confidence levels (often 95%).
Properties:

- perfect linearity shows perfect fit of F to the data, while strong curvature suggests poor fit;
- outliers show as extreme values lying well off the line of the other data;
- for standard normal Q–Q plots we use $x_i = \Phi^{-1}\left(\frac{i}{n+1}\right)$, where Φ is the $\mathcal{N}(0,1)$ distribution function.

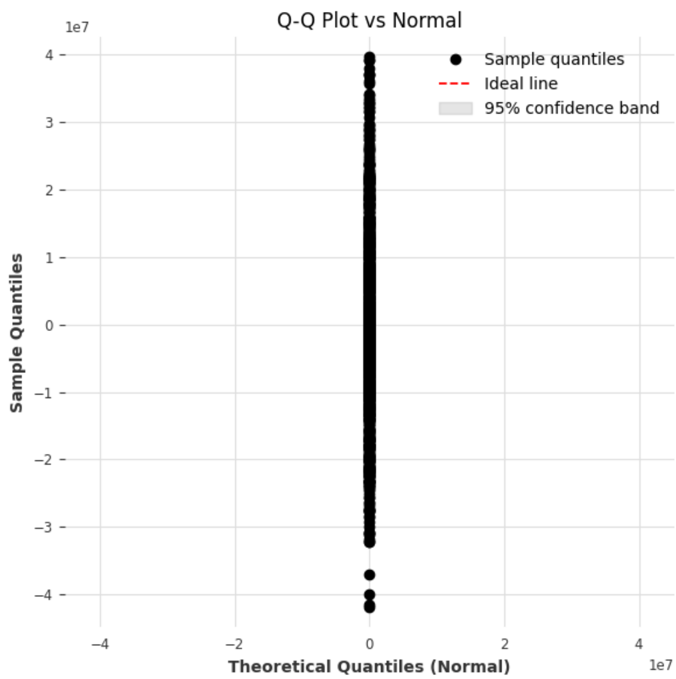


Figure 24: Enter Caption

Standardization The graph is too close together, to have a clearere view , we must standarize the data before generating the Q–Q plot. we do so using:

$$z_i = \frac{x_i - \bar{x}}{s}$$

where $\bar{x}$ and s are the sample mean and standard deviation, respectively.

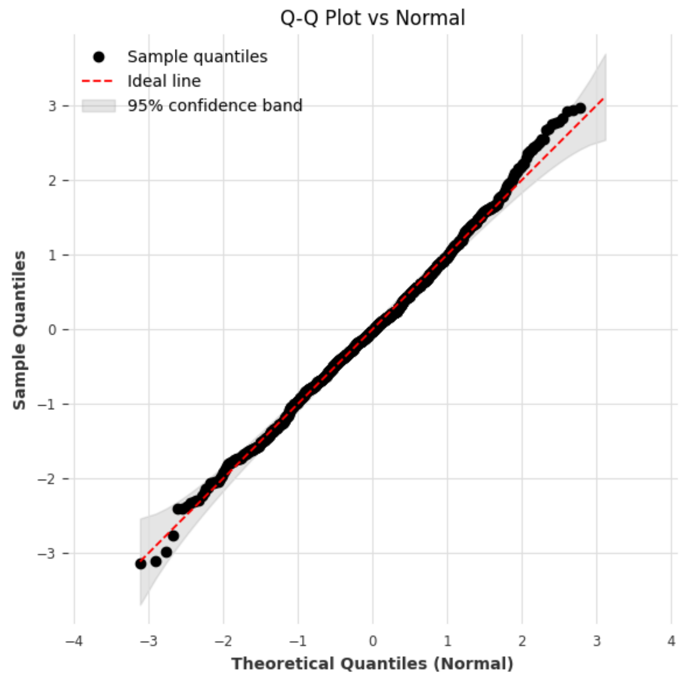


Figure 25: Enter Caption

4.5 Periodogram

4.5.1 Motivation

Many series have cyclic structure (e.g. sunspots, CO2 data,...), but we may not know what the cycles are in advance of looking at the data. The periodogram is a summary description based on representing the observed series as a superposition of sine and cosine waves of various frequencies. [4]

Check si il y a plusieurs fréquences, lequels, c'est quoi leurs amplitudes, check residual stat

4.5.2 Discrete Fourier transform

We can avoid the previous regression and use the discrete Fourier transform (DFT) for frequency analysis of time series.

The discrete Fourier transform of a time series $y_1, \ldots, y_n$ is the complex-valued series

$$d(\omega_j) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} y_t e^{-2\pi i \omega_j t}$$

$$d(\omega_j) = \frac{1}{\sqrt{n}} \left(\sum_{t=1}^{n} y_t \cos(2\pi\omega_j t) - i \sum_{t=1}^{n} y_t \sin(2\pi\omega_j t) \right)$$

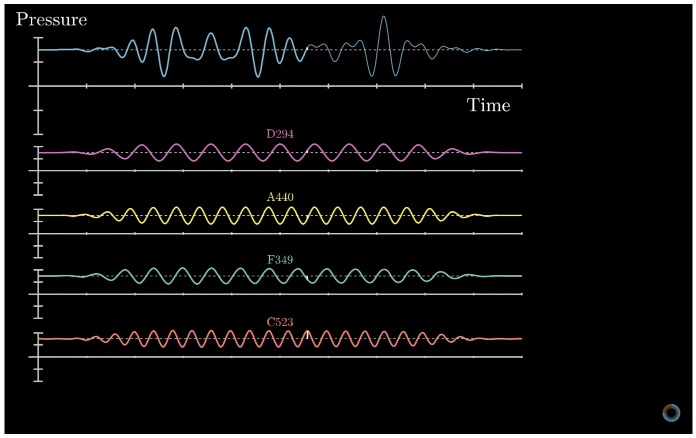


Figure 26: Signal decomposition using Fourier Transform (3Blue1Brown) [13]

We define the periodogram $I(\omega_j) = |d(\omega_j)|^2$.
The periodogram is related to the scaled periodogram: $I(\omega_j) = \frac{n}{4} P(\omega_j)$.

4.5.3 Periodogram

(a) If $y_1, \ldots, y_n$ is an equally-spaced time series, its periodogram ordinate for ω is defined as

$$I(\omega) = |d(\omega_j)|^2$$

this means that:

$$I(\omega) = \frac{1}{n} \left[\left(\sum_{t=1}^{n} y_t \cos(2\pi\omega t) \right)^2 + \left(\sum_{t=1}^{n} y_t \sin(2\pi\omega t) \right)^2 \right], \quad 0 < \omega \leq \frac{1}{2}$$

Our plot for the linear periodogram:

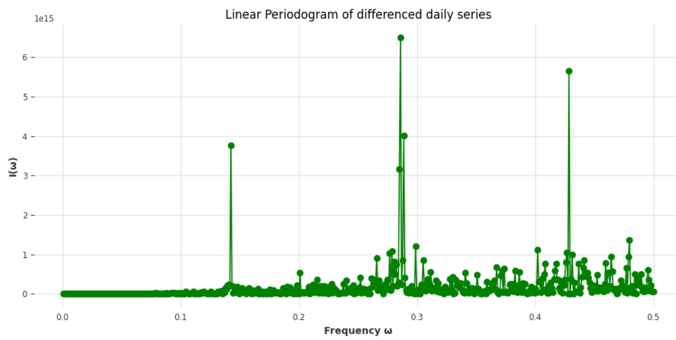


Figure 27: Enter Caption

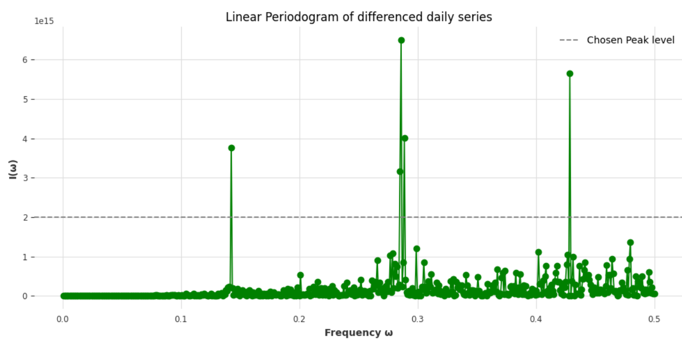


Figure 28: Enter Caption

4.5.4 Spectral Analysis— Power Spectrum

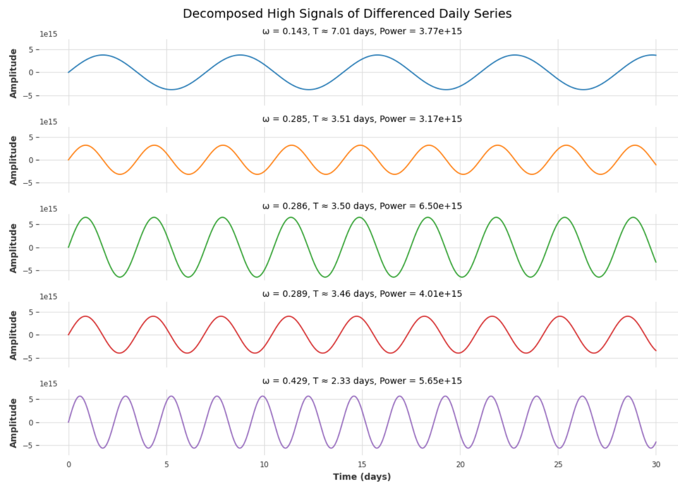


Figure 29: Enter Caption

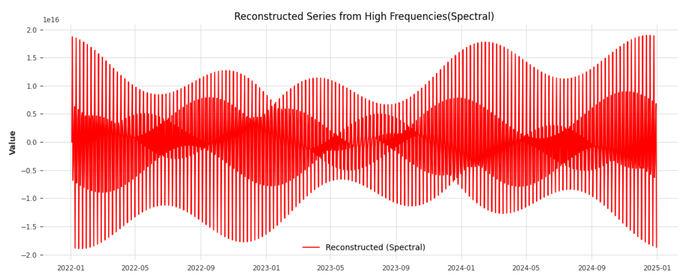


Figure 30: Enter Caption

4.5.5 Cumulative periodogram

(c) The cumulative periodogram

$$C_r = \frac{\sum_{j=1}^{r} I(\omega_j)}{\sum_{l=1}^{m} I(\omega_l)}, \quad r = 1, \ldots, m$$

is a plot of $C_1, \ldots, C_m$ against the frequencies ω_j for $j = 1, \ldots, m$. [4]

According to Davidson, Gaussian and non-Gaussian white noise has a flat spectrum [4]

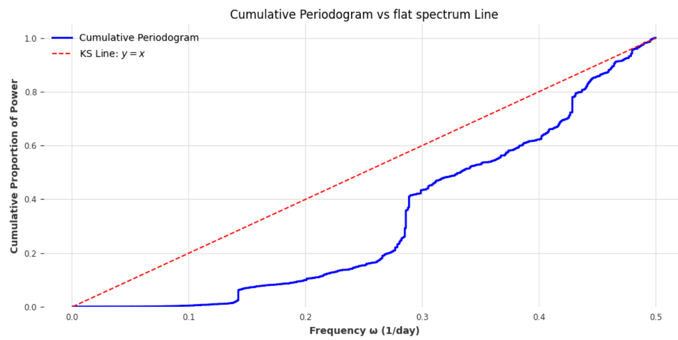


Figure 31: Enter Caption

4.5.6 Interpretation

Spectral analysis of variance

Might be added if time allows

4.6 Smoothing

Smoothing data set is to create an approximating function that attempts to capture important patterns in the data, while leaving out noise or other fine-scale structures. [15]

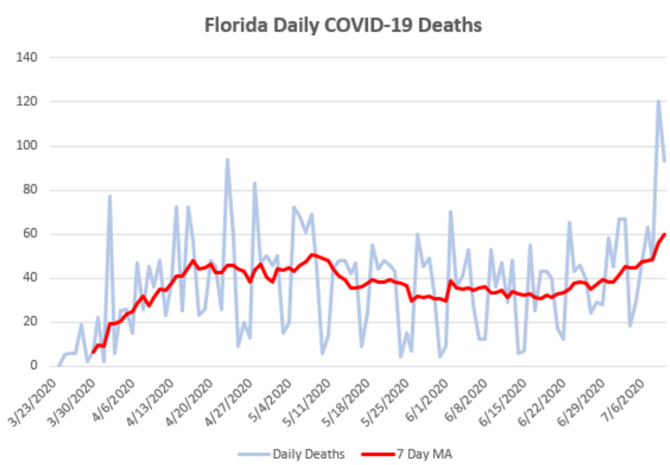


Figure 32: Smoothing Example - COVID Deaths 2019 (statisticsbyjim.com) [8]

4.6.1 Motivation

According to Boris's book, the underlying model is

$$Y_t = \mu(t) + Z_t,$$

[4] where $\mu(t)$ is smooth function of t and $\{Z_t\}$ is stationary. Among other things, smoothing can identify trends and seasonality. Differencing can remove trend to give stationary series. But differencing does not allow us to visualise the trend. [4]

We can implement smoothing to examine/estimate the trend, for example using:

- moving average (simple, related to differencing);
- polynomial (simple, doesn't work very well);
- local polynomial (simple, easy to robustify);
- STL decomposition (robust fitting of local polynomial, with seasonal effects). [4]

We will see later that using differencing results in large uncertainties in predictions. Intuitively, this is because differencing can remove very random trends which must be taken into account for later predictions of Y_{t+h}. If we can estimate the trend $\mu(t)$ accurately and predict it with low uncertainty, we can obtain better forecasts than when using differencing. [4]

4.6.2 Moving averages

Simple Moving Average

Moving Averages is one of the simplest smoothing methods to implement, it essentially computes the local average function of a window size = n.

Simple Moving Average formula is given by the formula:

$$\text{SMA}_t = \frac{1}{N} \sum_{i=0}^{N-1} y_{t-i}$$

[14]

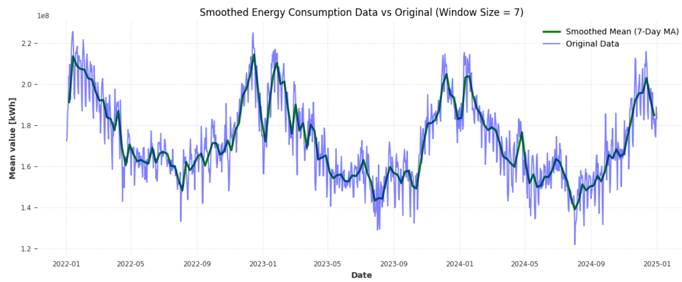


Figure 33: Enter Caption

Weighted Moving average

Weighted Moving average is given by the formula :

$$s_t = \sum_{j=-p}^{p} w_j y_{t+j}, \quad t = p+1, \dots, n-p, \quad p \in N,$$

[4]

Classical approach to smoothing: given data $y_1, \dots, y_n$, replace y_t by $\frac{1}{3}(y_{t+1} + y_t + y_{t-1})$, or in general construct the moving average of order $2p+1$, and weights w_j, with $\sum w_j = 1$ and (usually) $w_j > 0$ and $w_j = w_{-j}$. This is an example of a linear filter.

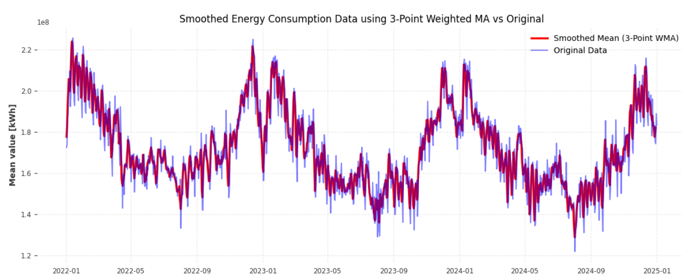


Figure 34: Smoothed Energy Consumption Data using 3-Point Weighted MA vs Original

[4]

Fixing Weights

Fixes are possible near the ends, but usually $p \ll n$, so the details of the fixes are unimportant. Choose weights by:

- iterating simple (equally-weighted) smoothers;

- choosing higher order to remove (or at least decrease) seasonality, for example taking $p = 6$, $w_6 = w_{-6} = 1/24$ and all other $w_j = 1/12$;
- taking smaller order to highlight seasonality. [4]

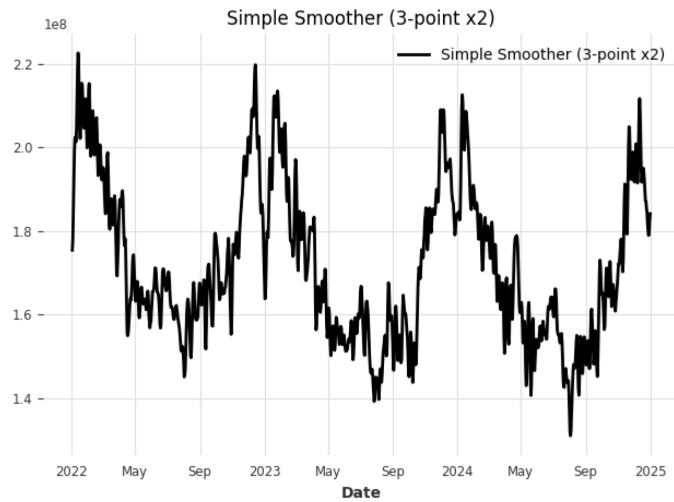


Figure 35: Higher Order Smoother, weights 1/24 for w1, w2p+1, 1/12 for middle weights

4.6.3 Local polynomial regression

A gloabl fit polynomial of degree k to the data is given by the formula

$$Y_t = p(t) + \varepsilon_t = \beta_0 + \beta_1 t + \cdots + \beta_k t^k + Z_t,$$

where $\{Z_t\}$ is stationary series. Choose parameters. [4] $\beta_0, \ldots, \beta_k$ to minimise the sum of squares

$$\sum_{t=1}^{n} \{y_t - p(t)\}^2 = \sum_{t=1}^{n} \left[y_t - \left(\beta_0 + \beta_1 t + \cdots + \beta_k t^k\right)\right]^2,$$

[4]

Instead of fitting one global curve , we fit small polynomial curves locally, near each time point. We give more weight to nearby points and less weight to far-away ones using a "Kernel function" that assigns these weights, where we pick a time point, and assign weights to nearby observations using the kernel function, and repeat for the next time point.

Automatic choice of h (or equivalent degrees of freedom $\equiv$ degree of polynomial) for kernel tends to be too small, owing to autocorrelation of time series. [4]

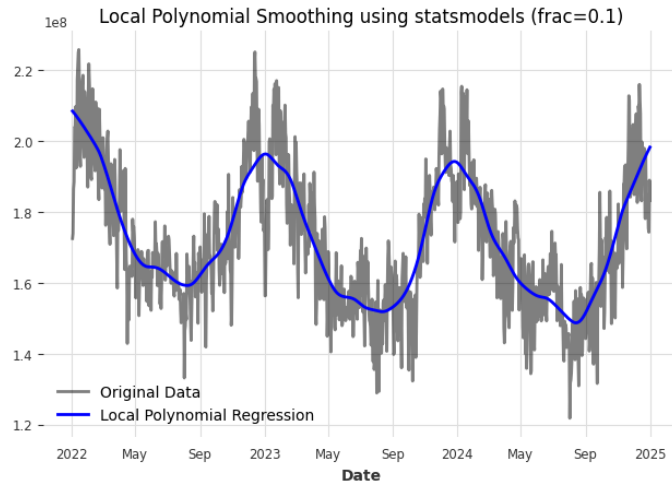


Figure 36: Local polynomial regression using statsmodels (p=0.25)

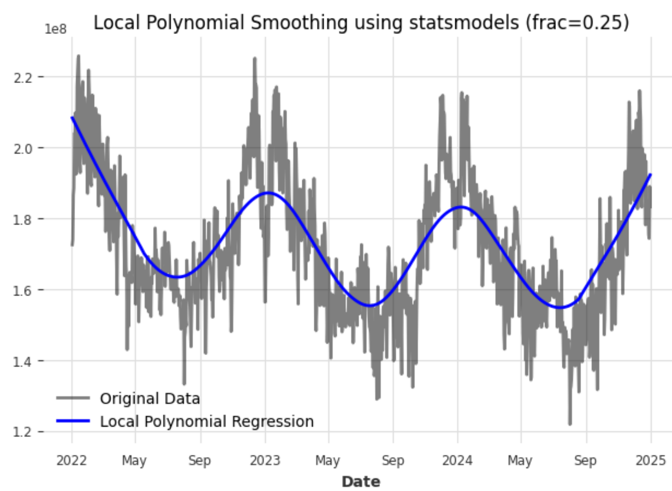


Figure 37: Local polynomial regression using statsmodels (p=0.25)

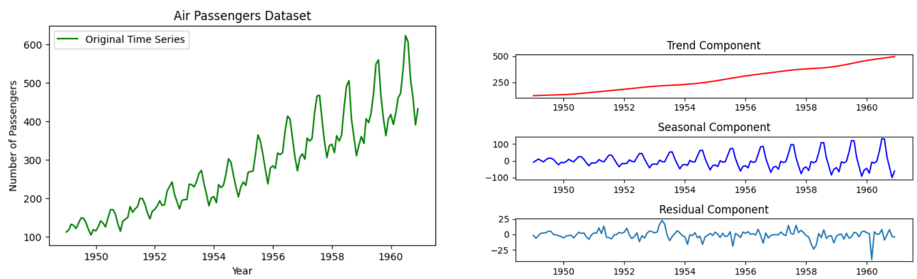
4.6.4 STL decomposition

An approach to removing overall trend and seasonal components, robust and (in principle) capable of handling missing data. [4]

The underlying model is:

$$Y_t = U(t) + S(t) + Z_t, \quad \{Z_t\} \text{ stationary,}$$

where $U(t)$ is the trend component and $S(t)$ is the seasonal variation. [4]



(a) Original data [10]

(b) STL decomposition [10]

Figure 38: Example STL decomposition: (geeksforgeeks) [10]

Method: STL decomposition is based on the local polynomial regression. The seasonal component is found by smoothing the seasonal sub-series. This seasonal component is then removed from the initial data, and the remainder is smoothed to estimate the trend component. [4]

STL can fit either a single seasonal component or a slowly-varying one. There are several parameters to be chosen when using STL. The default values in the `stl` function are not always appropriate. [4]

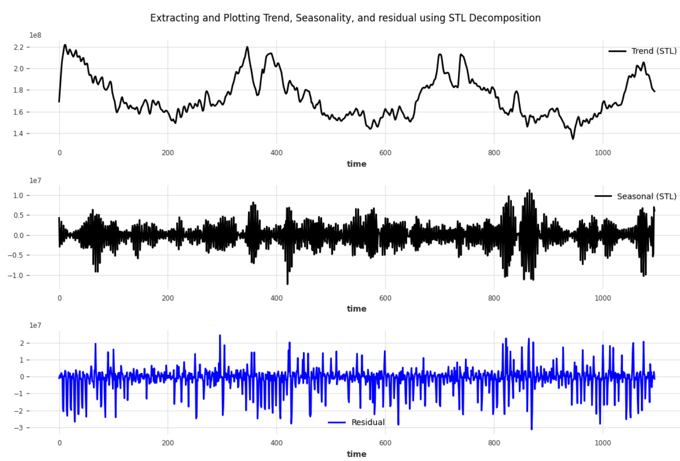


Figure 39: STL extracting trend seasonality

5 Additive vs Multiplicative Extraction Models

According to Mulius' article [12], in an additive model, the observed value is the sum of trend, seasonality, and residual components:

$$Y(t) = \text{Trend}(t) + \text{Seasonality}(t) + \text{Residual}(t).$$

This works best when the magnitude of seasonal fluctuations or residuals remains constant over time, regardless of the trend's growth or decline.

A multiplicative model represents the observed value as the product of its components:

$$Y(t) = \text{Trend}(t) \times \text{Seasonality}(t) \times \text{Residual}(t).$$

This is suitable when seasonal or residual effects scale with the trend [12].

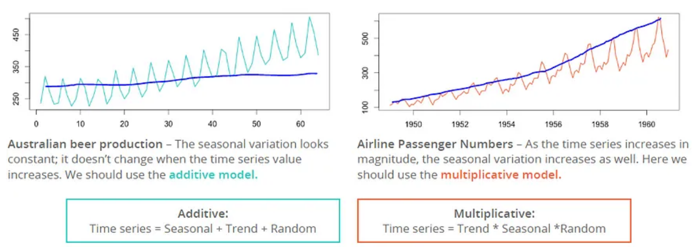


Figure 40: decomposition additive vs multiplicative (Nachi Keta) [11]

5.1 Models

5.1.1 Baseline Model

My baseline model will be a naïve forecast model. It assumes that the value at time t is equal to the value at the previous time step:

$$Y_t = Y_{t-1}$$

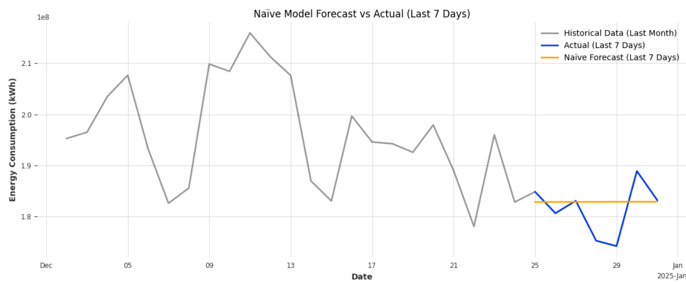


Figure 41: Baseline Naive model, 1 week daily forecast

5.1.2 AR(1)

The first-order autoregressive, AR(1), process is a stationary process $\{Y_t\}$ satisfying

$$Y_t = \alpha Y_{t-1} + \varepsilon_t, \quad t \in Z, \tag{1}$$

where α is the autoregressive parameter and $\{\varepsilon_t\}$ is white noise. The AR(1) process with mean μ is defined by

$$Y_t - \mu = \alpha(Y_{t-1} - \mu) + \varepsilon_t, \quad t \in Z. \tag{2}$$

In theoretical discussion, we use (1), but in practice we must usually use (2) and estimate the mean μ.[4]

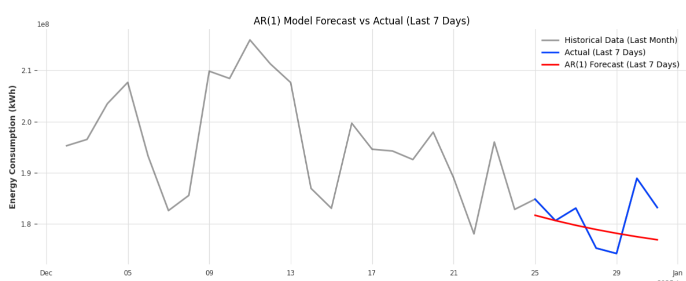


Figure 42: AR(1,0,0) model, 1 week daily forecast

5.1.3 ARIMA

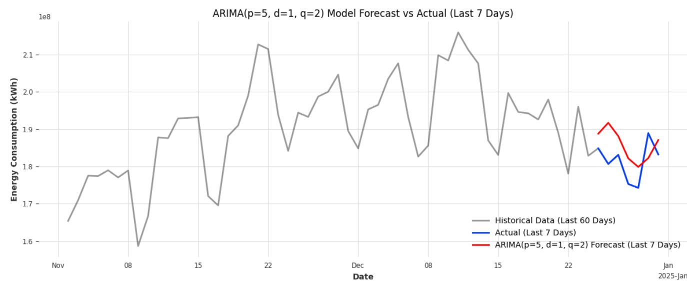


Figure 43: Enter Caption

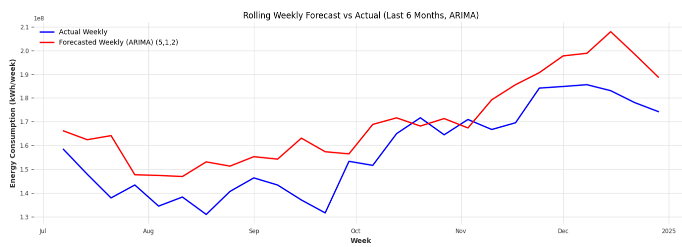


Figure 44: Enter Caption

5.1.4 SARIMA

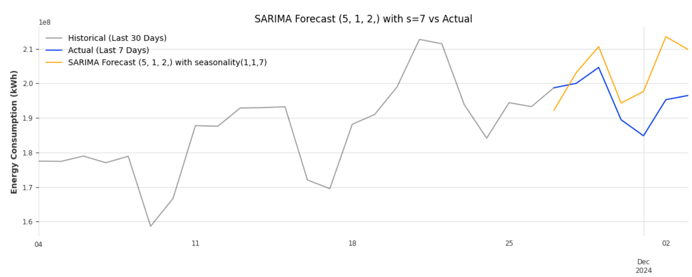


Figure 45: Enter Caption

5.1.5 Regularization Techniques

5.2 Evaluation

References

[1] André Bauer. *Automated Hybrid Time Series Forecasting: Design, Benchmarking, and Use Cases.* Doctoral thesis, Illinois Institute of Technology, 2021. License: CC BY-SA 4.0.

[2] Wikipedia contributors. Mean – Wikipedia, the free encyclopedia, 2024. Accessed: 2024-05-21.

[3] Wikipedia contributors. Variance – Wikipedia, the free encyclopedia, 2024. Accessed: 2024-05-21.

[4] Anthony Davison and Emeric Thibaud. *Time Series.* EPFL, 2019. MATH-342 Course, Anthony Davison © 2019.

[5] Figure 1. Electrical engineering portal. `https://engineering.electrical-equipment.org`.

[6] Figure 2. Coded using swissgrid data. Custom dataset, not published.

[7] Figure 3. Grid 2040 – nepal electricity authority. `https://grid2040.ku.edu.np`.

[8] Jim Frost. Using moving averages to smooth time series data, 2024. Accessed: 2025-05-30.

[9] GeeksforGeeks. What is a trend in time series? `https://www.geeksforgeeks.org/what-is-a-trend-in-time-series/`. Accessed: 2025-05-21.

[10] GeeksforGeeks Contributors. Seasonal decomposition of time series by loess (stl), 2021. Accessed: 2025-05-30.

[11] Nachi Keta. Do you know in what context the terms additive and multiplicative are used in the context of time series? `https://nachi-keta.medium.com/do-you-know-in-what-context-the-terms-additive-and-multiplicative-are-used-in-the-cont` 2021. Accessed: 2025-05-30.

[12] Milvus AI Contributors. What is the difference between additive and multiplicative time series models? `https://milvus.io/ai-quick-reference/what-is-the-difference-between-additive-and-multiplicative-time-series-models`, 2024. Accessed: 2025-05-30.

[13] Grant Sanderson. But what is the fourier transform? a visual introduction. `https://www.youtube.com/watch?v=spUNpyF58BY`, 2018. 3Blue1Brown, YouTube.

[14] Wikipedia contributors. Moving average – wikipedia, the free encyclopedia, 2024. Accessed: 2025-05-30.

[15] Wikipedia contributors. Smoothing – wikipedia, the free encyclopedia, 2024. Accessed: 2025-05-30.

[16] Wikipedia contributors. White noise — wikipedia, the free encyclopedia. https://en.wikipedia.org/wiki/White_noise, 2024. Accessed: 2025-05-21.