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.

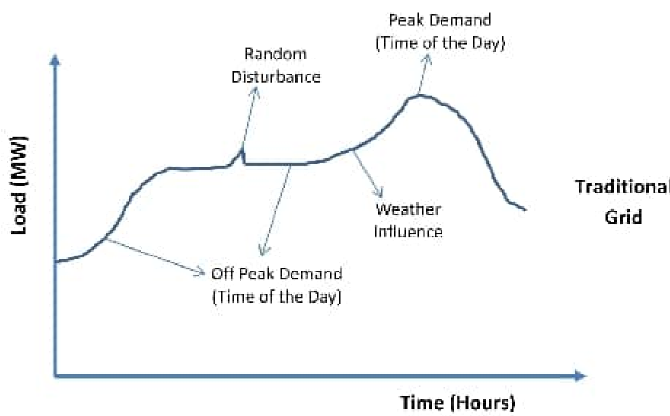


Figure 1: Load and influence

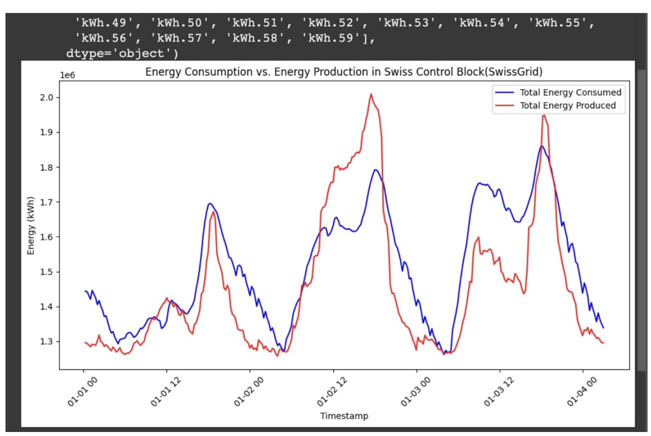


Figure 2: Plotted Energy Production visuals (Python3)

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

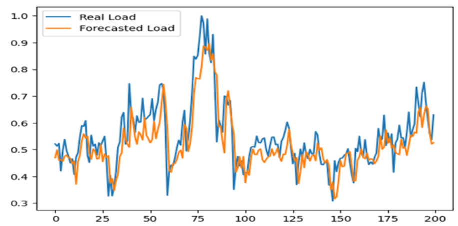


Figure 3: Endgoal

2 Objectives

The primary objectives are:

1. Visualize the Data
2. Develop a Medium Term Forecasting Model (Predicting Total Amount of Energy Consumed per 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

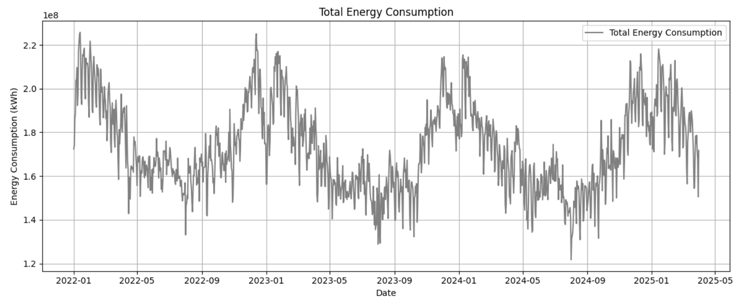


Figure 4: Total Swiss Energy Consumption(2024)

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]

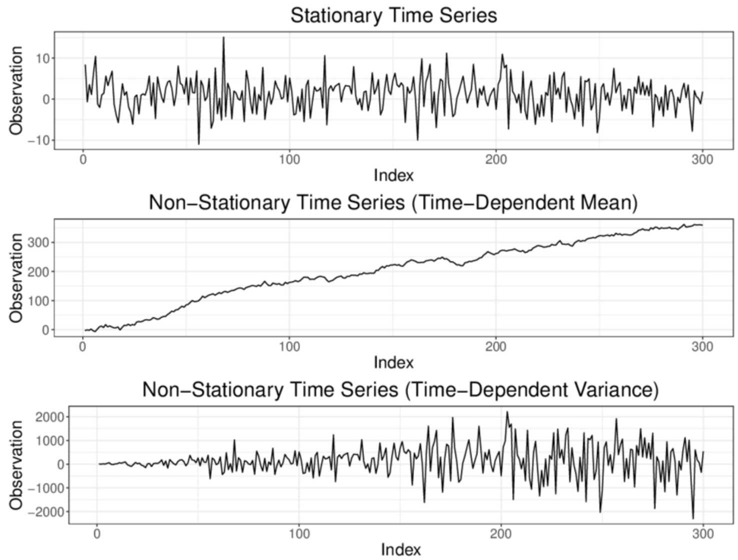


Figure 5: Stationary and non-stationary time series, Bauer 2021 [1]

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 [9]:

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

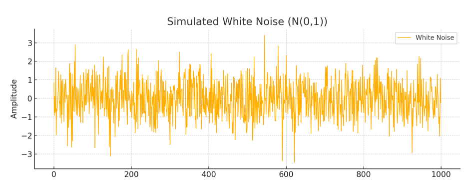


Figure 6: White noise using an np.random fct

As explained in Wikipedia [9], 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].

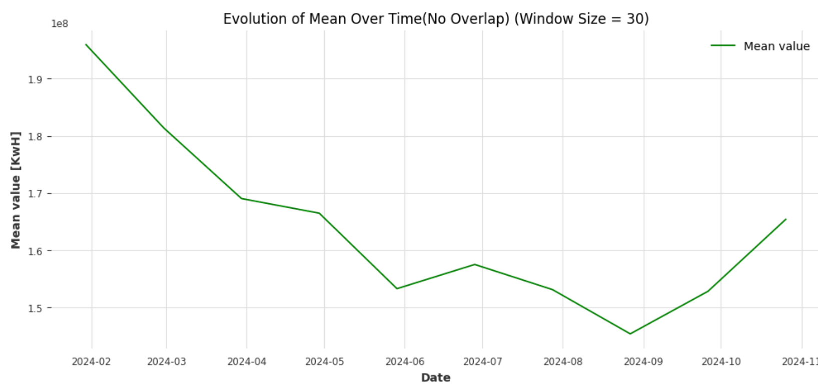


Figure 7: Evolution of Mean Over Time — 2024

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].

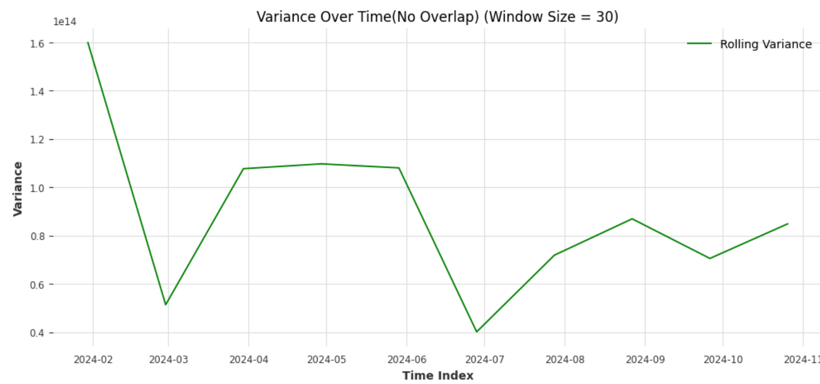


Figure 8: Evolution of Variance Over Time — Total Energy Consumption 2024

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.

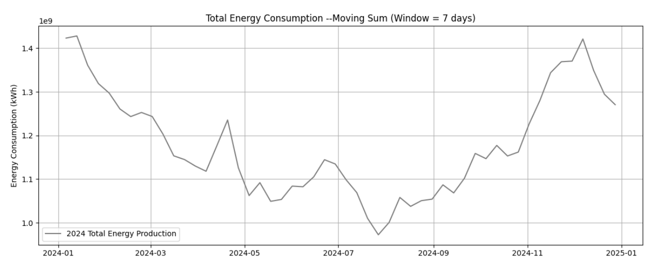


Figure 9: Weekly Sum

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].

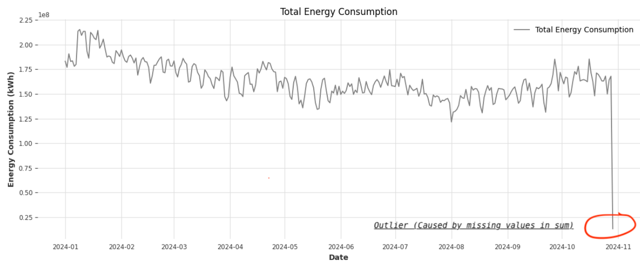


Figure 10: Outlier example in 2024 Consumption Graph

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 [8]. Trends can be linear, quadratic, periodic, or more complex [4].

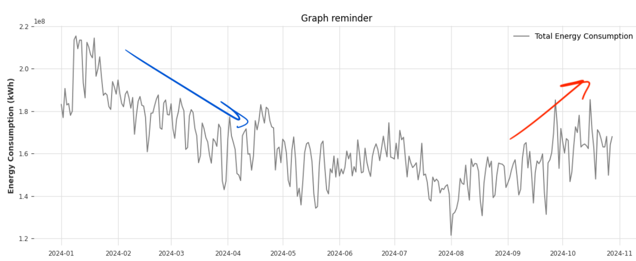


Figure 11: Trend Evolution in 2024 Consumption Graph

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).

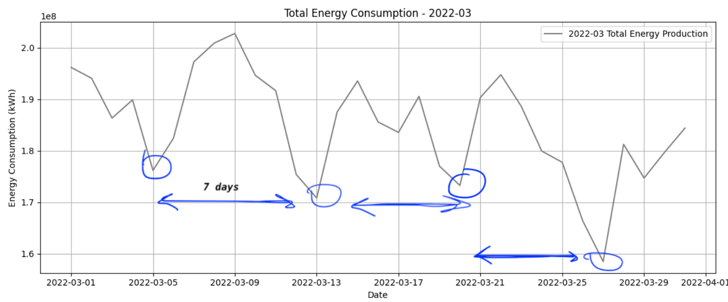


Figure 12: Weekly Seasonality, Month of March 2022

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]:

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

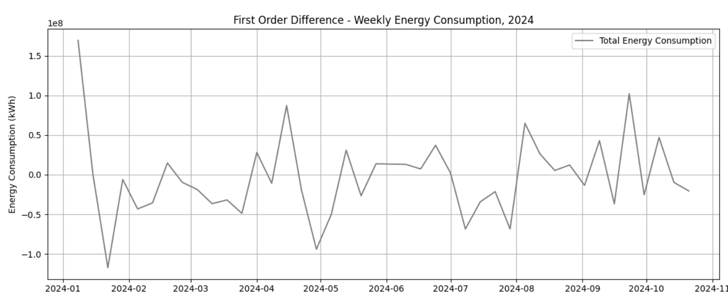


Figure 13: First Order Difference Weekly, 2024

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.

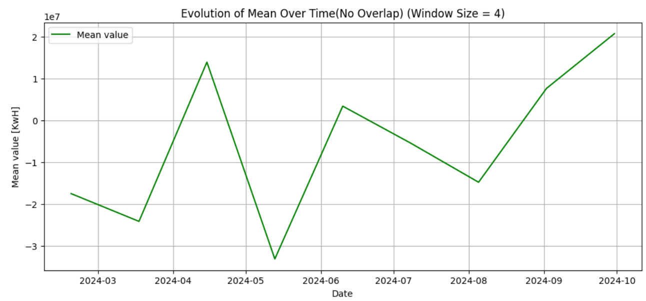


Figure 14: Mean Over Time after Differencing

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]:

$$\Delta^2 Y_t = \Delta Y_t - \Delta Y_{t-1} = Y_t - 2Y_{t-1} + Y_{t-2}$$

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].

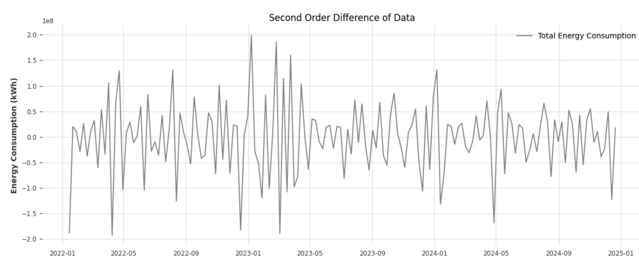


Figure 15: Second Order Difference Weekly, 2024

4.4 ACF and PACF

4.4.1 Motivation

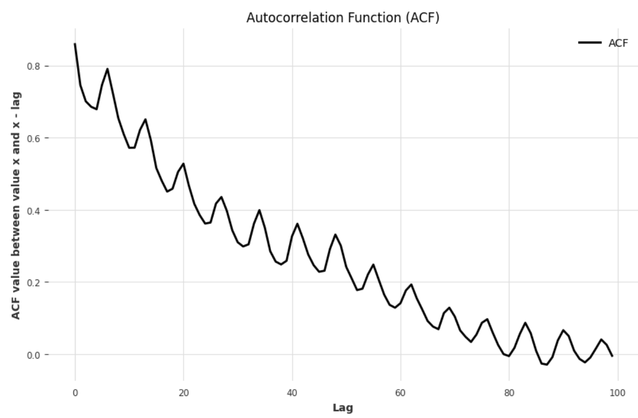
Autocorrelation and partial autocorrelation functions are used to understand the dependence structure of a time series. They help identify appropriate models [4].

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 $Y_{t-1}, \ldots, Y_{t-h+1}$.

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

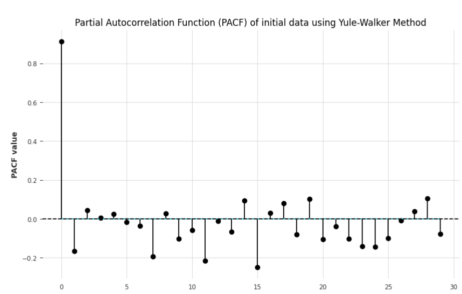


Figure 17: PACF computed using the Yule-Walker method

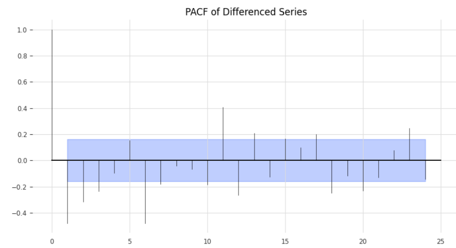


Figure 18: PACF generated using Darts

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.

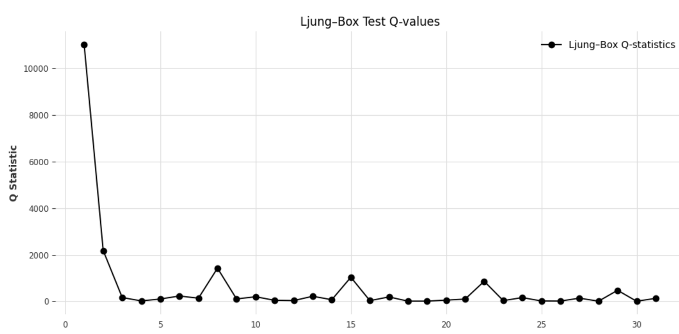


Figure 20: Enter Caption

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.

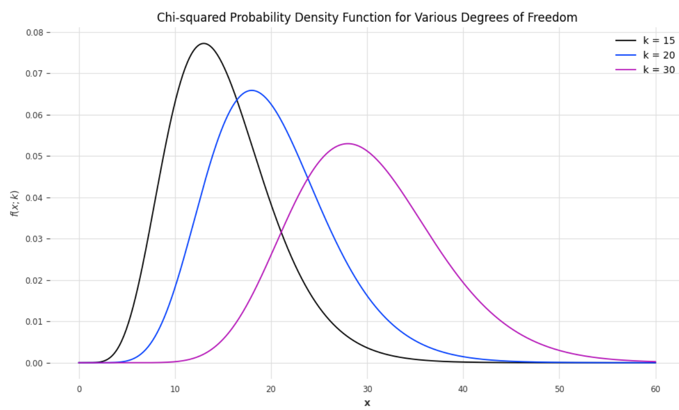


Figure 21: Enter Caption

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.

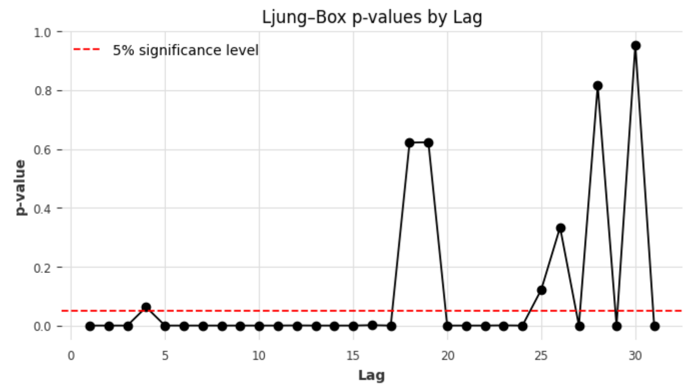


Figure 22: Enter Caption

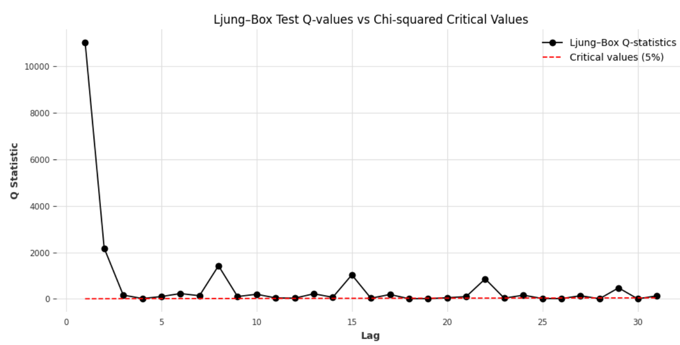


Figure 23: Enter Caption

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.

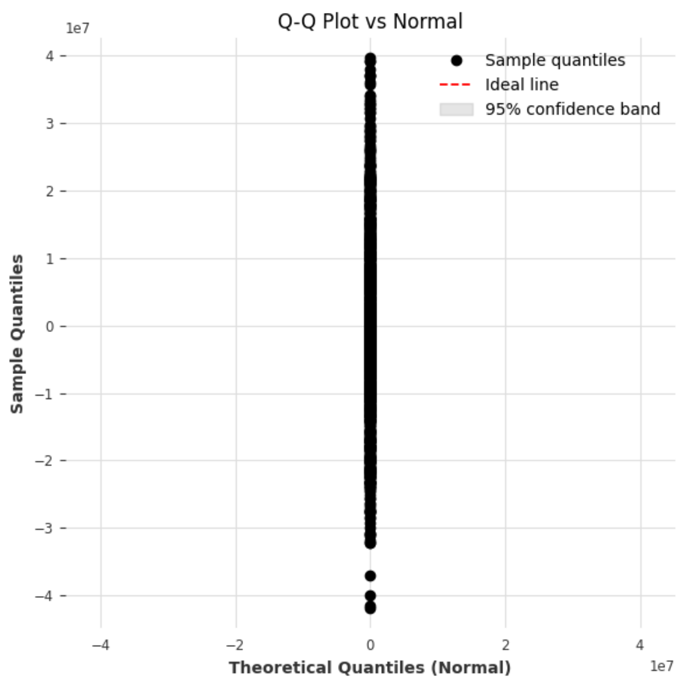


Figure 24: Enter Caption

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.

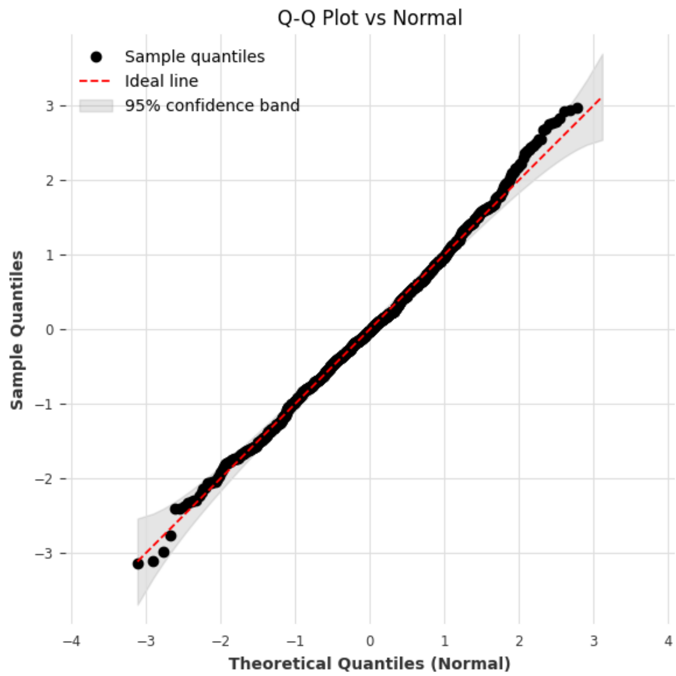


Figure 25: Enter Caption

4.5 Periodogram

4.5.1 Motivation

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

4.5.2 Discrete Fourier transform

4.5.3 Periodogram

4.5.4 Spectral Analysis— Power Spectrum

4.5.5 Properties of the periodogram

4.6 Smoothing

4.6.1 Motivation

4.6.2 STL decomposition

4.6.3 Moving averages

4.6.4 Local polynomial regression

4.7 Models

4.7.1 Baseline Model

4.7.2 AR(1)

4.7.3 AR(2)

4.7.4 AR(3)

4.7.5 Regularization Techniques

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

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