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.

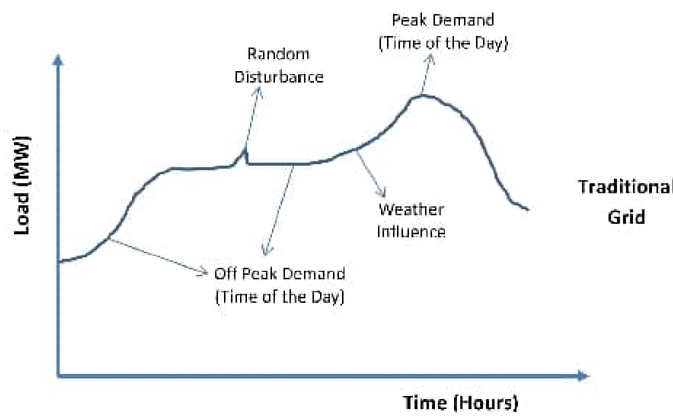

Figure 1: Load and influence

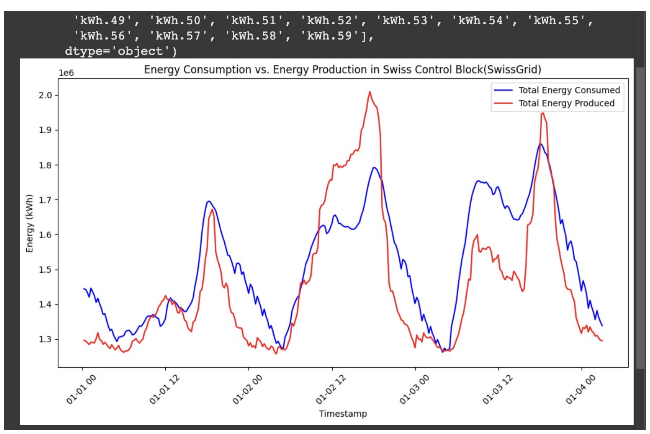

Figure 2: Plotted Energy Production visuals (Python3)

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

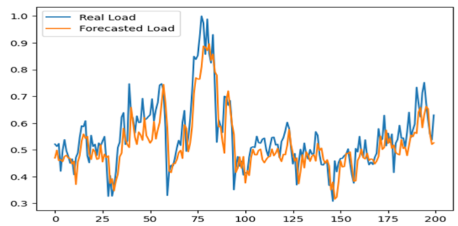

Figure 3: Endgoal

2 Objectives

The primary objectives are:

1. Develop a scalable, modular framework for load forecasting that can be integrated into real-time energy grid management systems.

2. Implement the framework for use in operational energy load prediction systems.

Visualisation

2.1 Variable to Predict

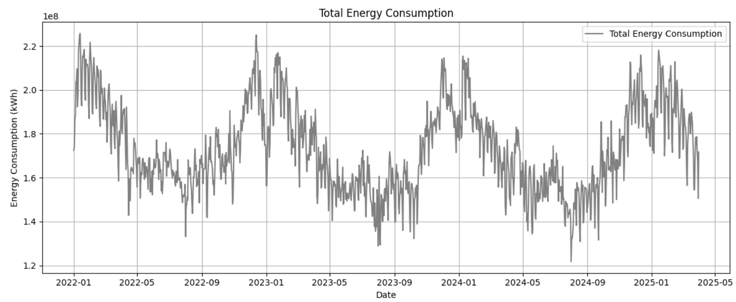

Figure 4: Total Swiss Energy Consumption(2024)

Data Cleaning

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

Time Series Analysis

2.2 Mathematical basics

2.2.1 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]

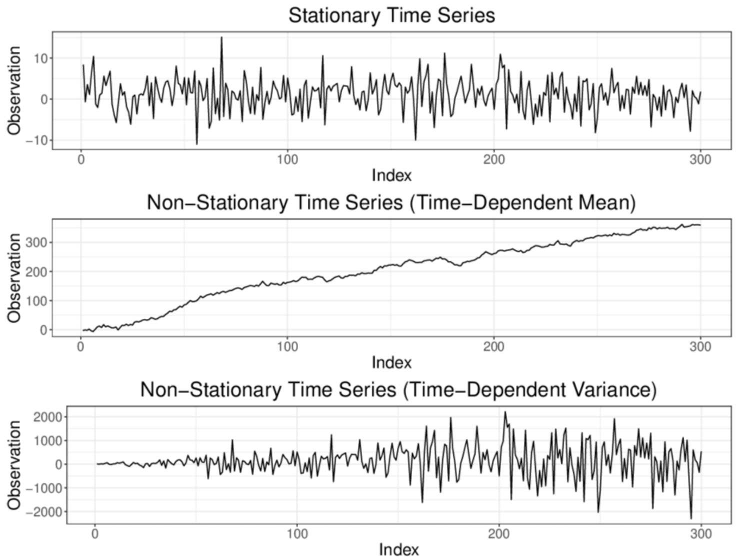

Figure 5: Stationary and non-stationary time series, Bauer 2021 [1]

2.2.2 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 $\mathrm{Var}(Y_t) = \sigma^2$ [4]. The standard deviation is measured by [9]:

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

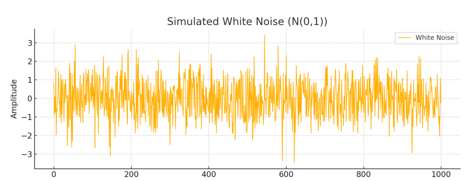

Figure 6: White noise using an np.random fct

As explained in Wikipedia [9], white noise is a stochastic process and does not have a deterministic function.

2.2.3 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]

2.3 Initial data analysis

2.3.1 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].

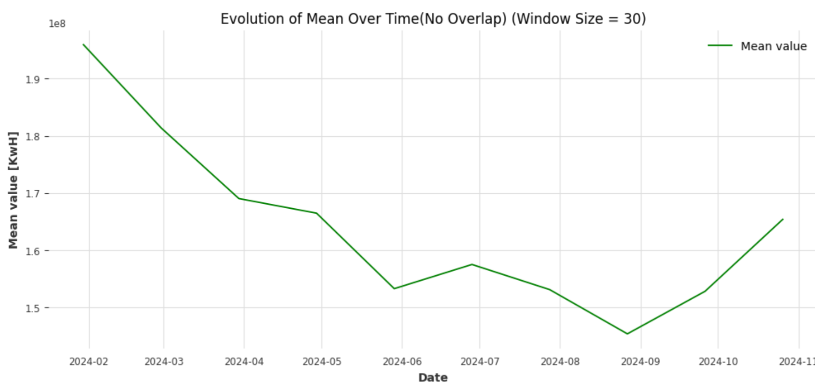

Figure 7: Evolution of Mean Over Time — 2024

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).

2.3.2 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].

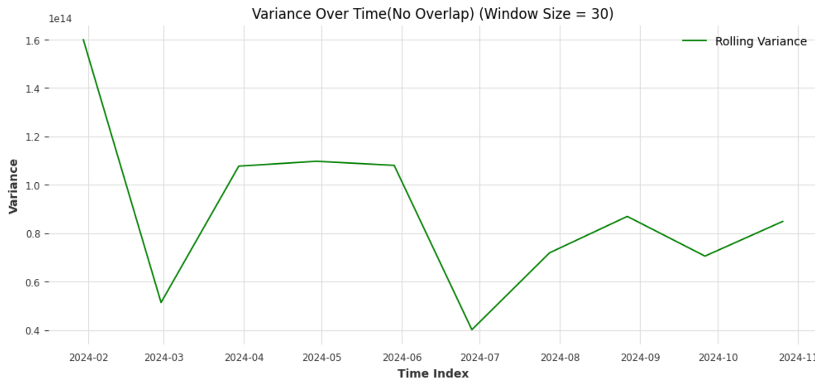

Figure 8: Evolution of Variance Over Time — Total Energy Consumption 2024

Observation: The variance fluctuates significantly over time.

2.3.3 Transformation

2.3.4 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.

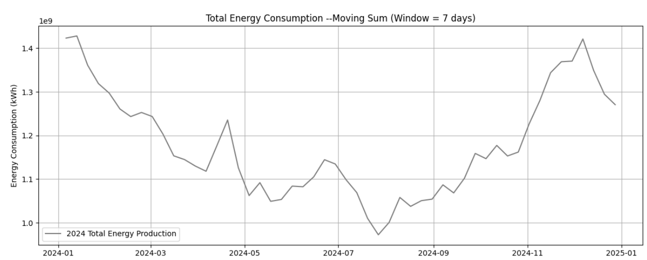

Figure 9: Weekly Sum

2.3.5 Moving Average

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

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

2.3.6 Outliers

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

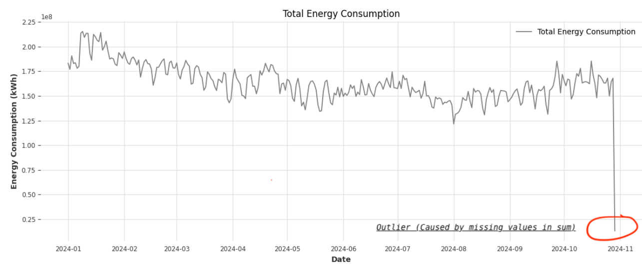

Figure 10: Outlier example in 2024 Consumption Graph

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

2.3.7 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].

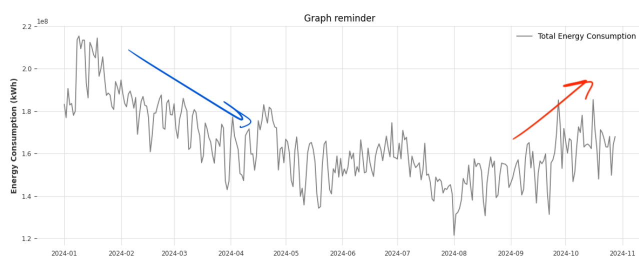

Figure 11: Trend Evolution in 2024 Consumption Graph

2.3.8 Seasonality

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

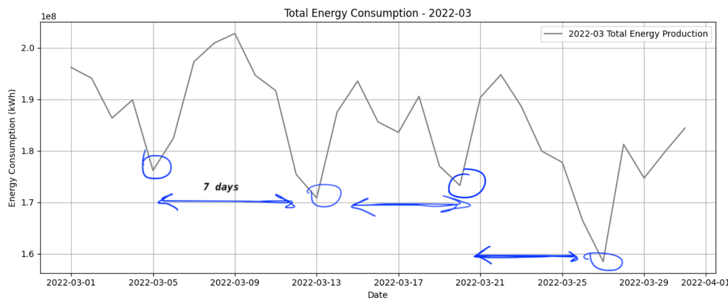

Figure 12: Weekly Seasonality, Month of March 2022

2.4 Differencing

2.4.1 Motivation

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

2.4.2 Differencing types

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

2.4.3 First-order difference

The first-order difference is defined as [4]:

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

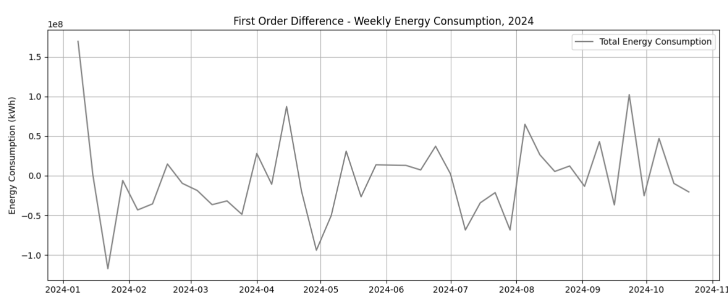

Figure 13: First Order Difference Weekly, 2024

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.

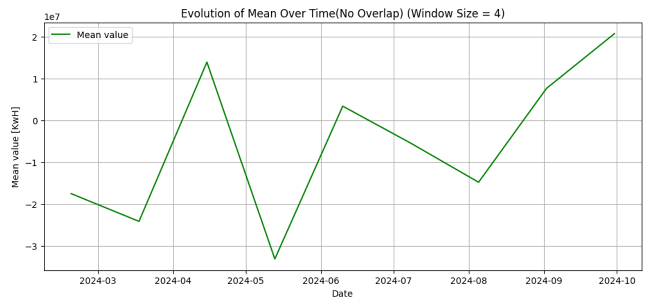

Figure 14: Mean Over Time after Differencing

2.4.4 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].

2.5 ACF and PACF

2.5.1 Motivation

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

2.5.2 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].

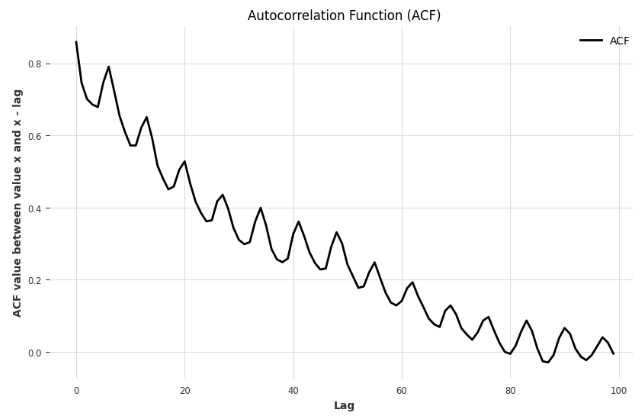

Figure 15: AutoCorrelation Function 2024 - Total Energy Consumption

2.5.3 Partial correlogram

The PACF is interpreted in the same way as the ACF. Let $Y_0, \ldots, Y_h$ be successive observations. The partial autocorrelation function (PACF) is:

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

2.5.4 Testing for stationarity

2.5.5 Testing for white noise

2.5.6 Checking normality using QQ plots

2.6 Periodogram

2.6.1 Motivation

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

2.6.2 Discrete Fourier transform

2.6.3 Periodogram

2.6.4 Properties of the periodogram

2.7 Smoothing

2.7.1 Motivation

2.7.2 STL decomposition

2.7.3 Moving averages

2.7.4 Local polynomial regression

2.8 Models

2.9 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.