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.

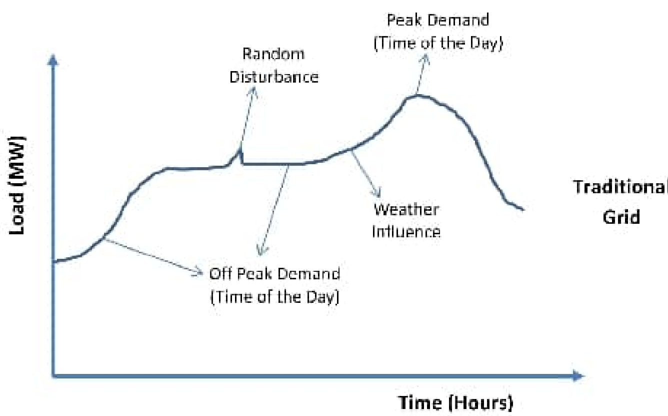


Figure 1: Load and influence

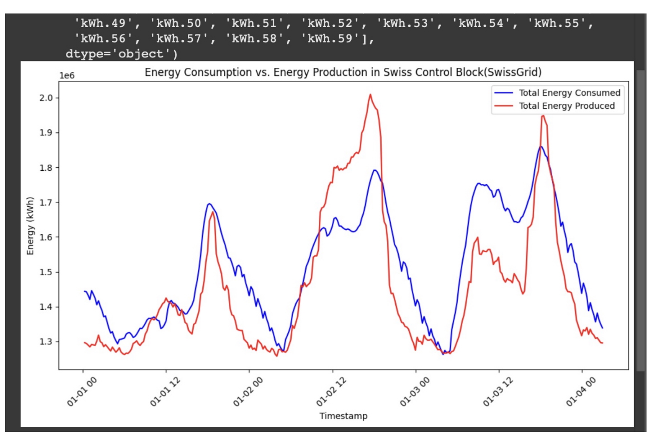


Figure 2: Plotted Energy Production visuals (Python3)

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

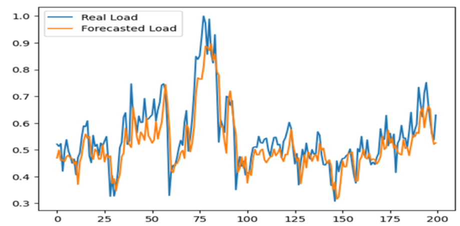


Figure 3: Endgoal

2 Objectives

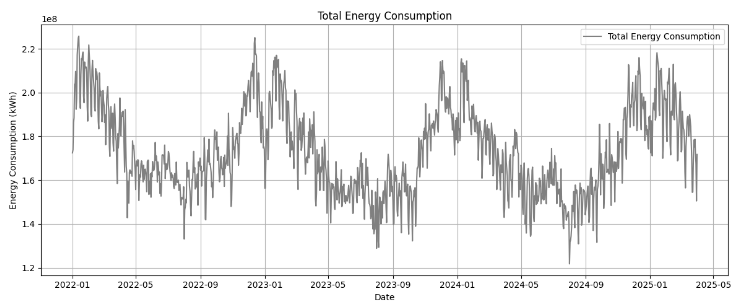
The primary objectives are:

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

The target variable I'm trying to predict is the weekly energy load, specifically **Total Energy Production**. The original data is recorded at 15-minute intervals, and I aggregate it to weekly values using a 7-day window (Monday to Sunday), following EU forecasting standards. According to Swissgrid, this variable represents the total energy produced in the control block Switzerland, based on aggregated feed-in sequences reported by distribution network operators. It includes only production plants equipped with load profile meters.



l

Figure 4: Total Swiss Energy Consumption(2024)

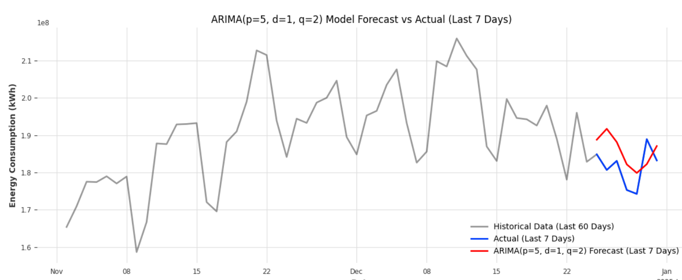


Figure 5: 1-week Forecasting example(ARIMA model)

3.2 Related Variables in the dataset

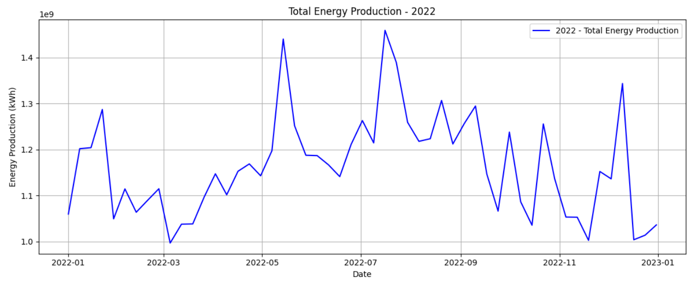


Figure 6: Total energy production over time

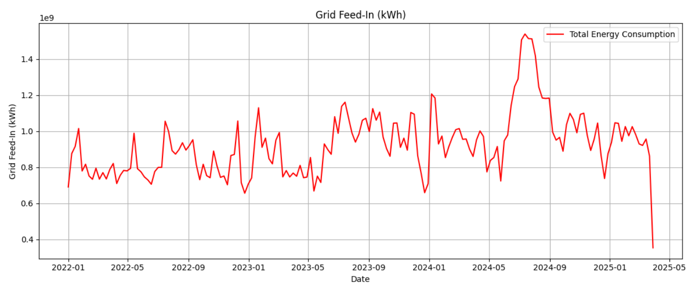


Figure 7: Electricity fed into the grid

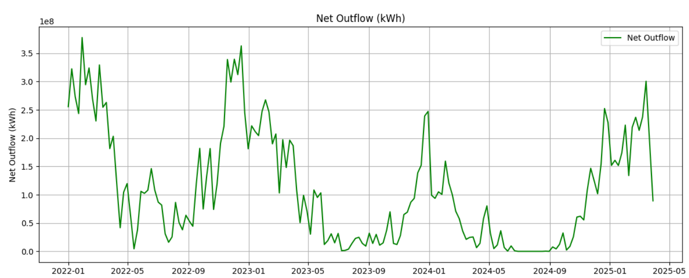


Figure 8: Net outflow of energy from the grid

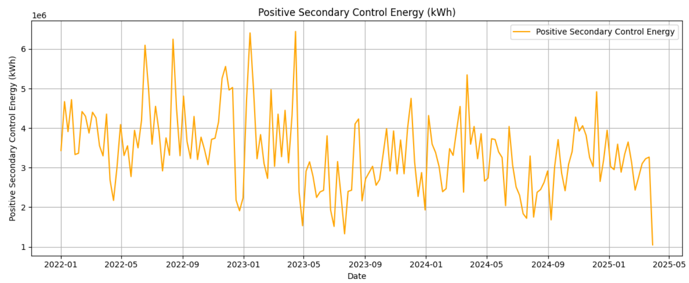


Figure 9: Activation of positive tertiary control reserves

4 Time Series Analysis

4.1 Data Cleaning

I started by converting everything to numeric, replacing errors with NaN, removing NaN rows

Images/Data Cleaning/empty_Loads_scatterplot.png

Figure 10: Scatterplot showing where the data had missing values before cleaning

For starters, no missing values were in the dataset (even before removing NaN rows).

Next, I checked if each timestamp in the dataset was unique. This is important for avoiding redundancy. Luckily, all indices turned out to be unique, so no extra work was needed.

Images/Data Cleaning/is_index_unique.png

Figure 11: Check showing that all index values are unique

I then looked at the summary stats (like mean, min, max, and standard deviation) of the energy consumption values. Nothing seems too off, which means there are probably no major outliers in the data.

Images/Data Cleaning/min_max.png

Figure 12: Stats summary: mean, min, max, and standard deviation of energy consumption

4.2 Mathematical basics

4.2.1 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]

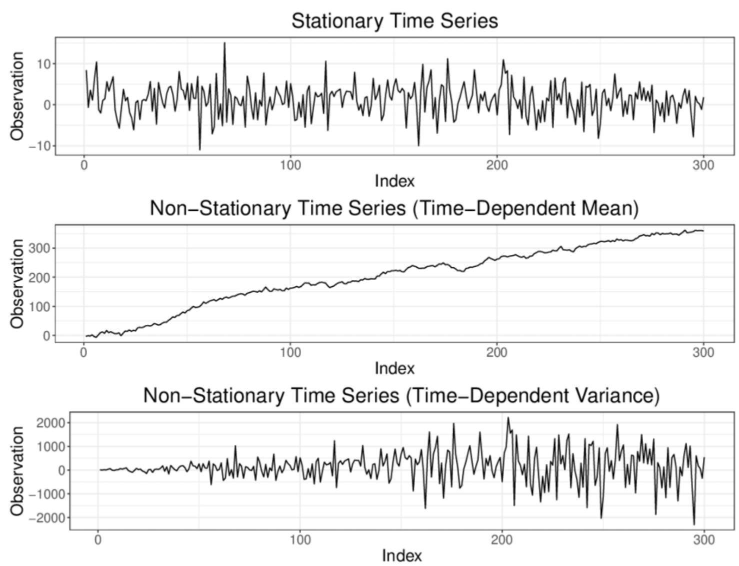


Figure 13: Stationary and non-stationary time series, Bauer 2021 [1]

4.2.2 White Noise

A stochastic process $\{Y_t\}$ is called *white noise* if all its elements are not correlated, with mean $\mathbb{E}(Y_t) = 0$ and constant variance $\mathrm{Var}(Y_t) = \sigma^2$ [4].

The standard deviation, which reflects the dispersion of values around the mean, is further explained in [11].

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

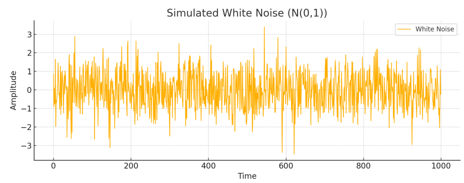


Figure 14: White noise using an np.random fct

As explained in Wikipedia [11], white noise is a stochastic process and does not have a deterministic function.

4.2.3 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]

4.3 Initial definitions

4.3.1 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].

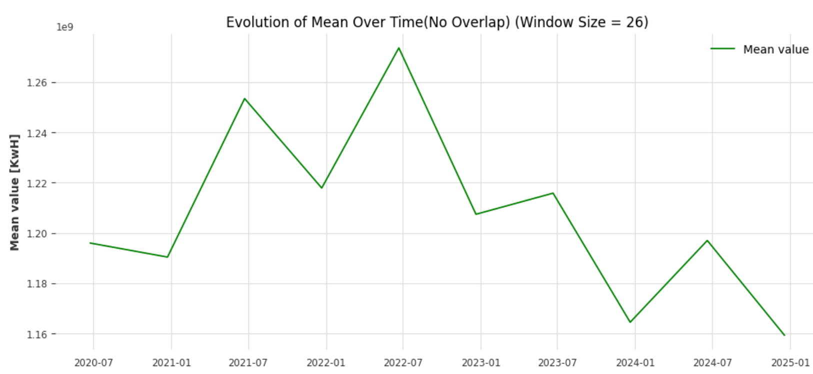


Figure 15: Evolution of Mean Over Time — 2024

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).

4.3.2 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].

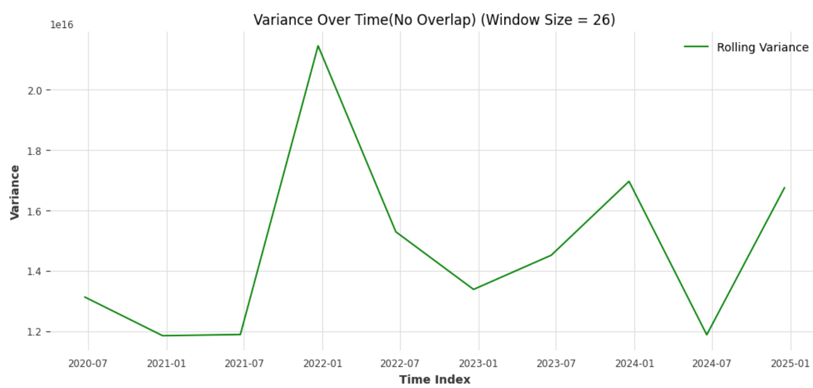


Figure 16: Evolution of Variance Over Time — Total Energy Consumption 2024

Observation: The variance fluctuates significantly over time.

4.3.3 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.

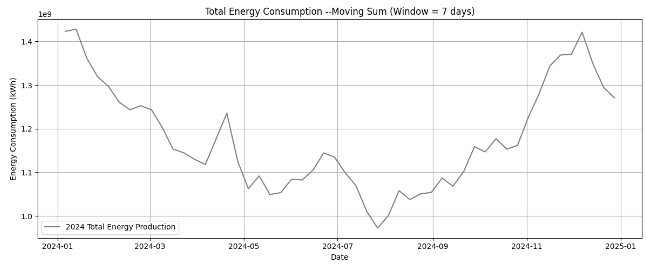


Figure 17: Weekly Sum

4.3.4 Moving Average

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

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

4.3.5 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].

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

4.3.6 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].

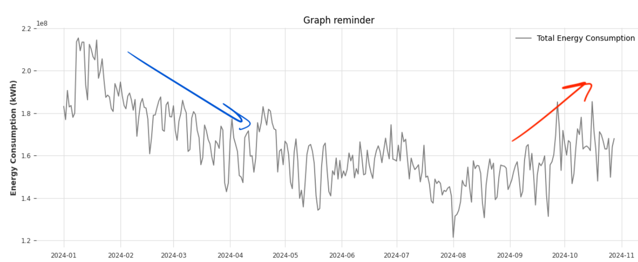


Figure 18: Trend Evolution in 2024 Consumption Graph

4.3.7 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).

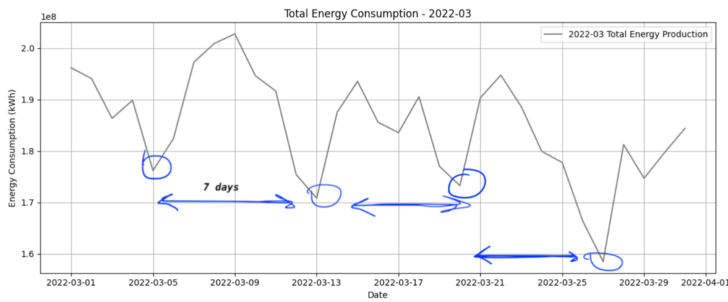


Figure 19: Weekly Seasonality, Month of March 2022

4.4 Differencing

4.4.1 Motivation

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

4.4.2 Differencing types

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

4.4.3 First-order difference

The first order difference is defined as [4]:

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

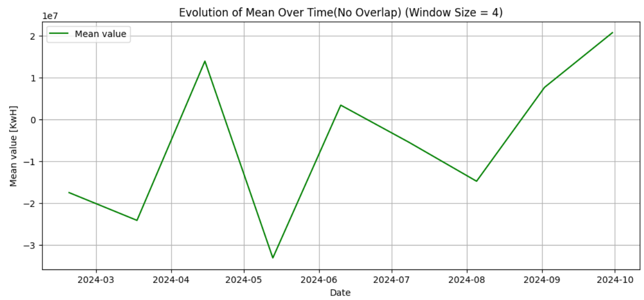


Figure 21: Mean Over Time after Differencing

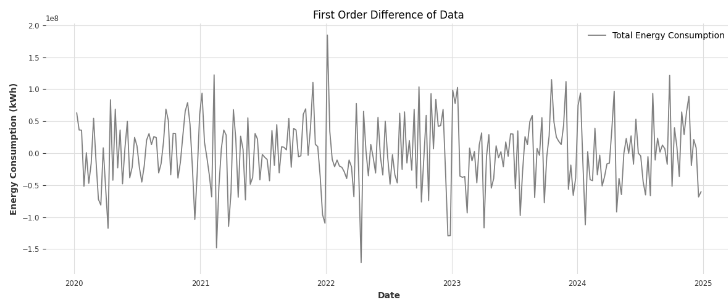


Figure 20: First Order Difference Weekly, 2024

Observation: The resulting time series does not look much different.

4.4.4 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]:

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

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].

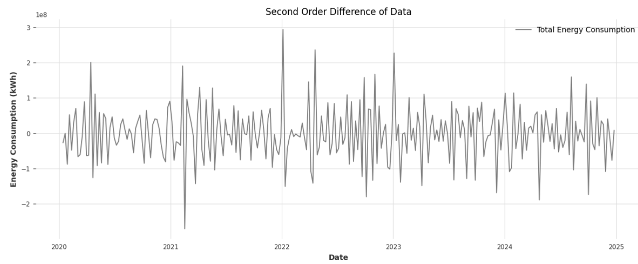


Figure 22: Second Order Difference Weekly, 2024

4.4.5 Seasonal differencing

It's simply (Yt - Yt - s), with s being the seasonality.

4.5 ACF

4.5.1 Motivation

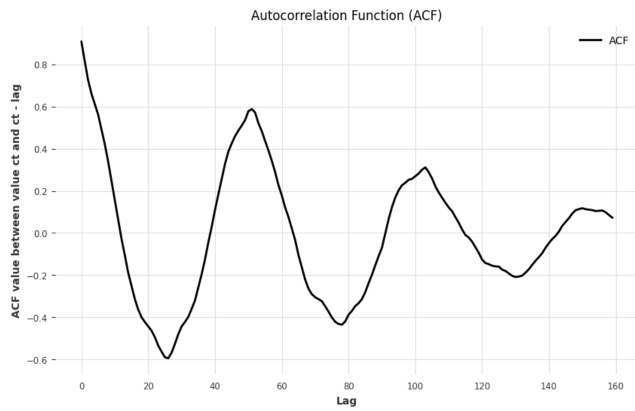
Autocorrelation and partial autocorrelation functions are used to understand the dependence structure of a time series. They help identify appropriate models [4].

4.5.2 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].



1

Figure 23: AutoCorrelation Function 2024 - Total Energy Consumption

4.5.3 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.

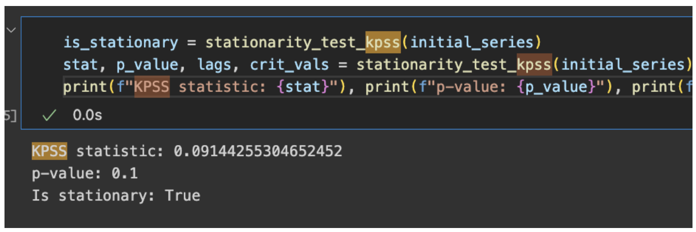


Figure 24: KPSS result for initial data

In this case, the KPSS test returns True for it is stationnary. This means that our time series is indeed stationnary.

4.6 Periodogram

4.6.1 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

4.6.2 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

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

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

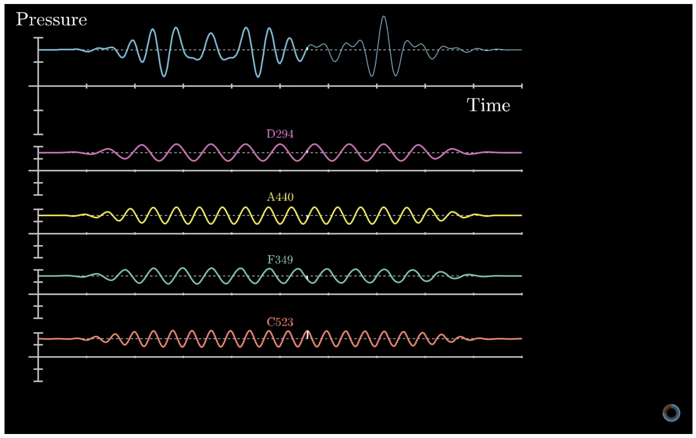


Figure 25: Signal decomposition using Fourier Transform (3Blue1Brown) [9]

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

4.6.3 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:

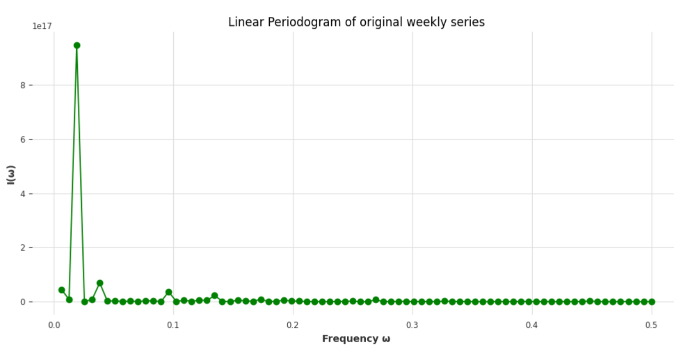


Figure 26: Linear Periodogram of original weekly series

4.6.4 Spectral Analysis— Power Spectrum

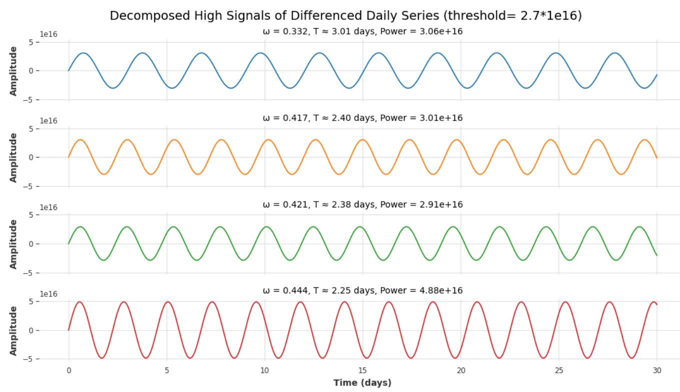


Figure 27: Decomposed Periodogram showing High Signals of Differenced Daily Series

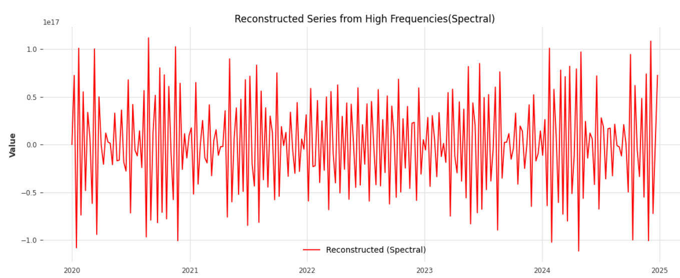


Figure 28: Spectral Reconstructed Series from High Frequencies(Spectral)

4.6.5 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]

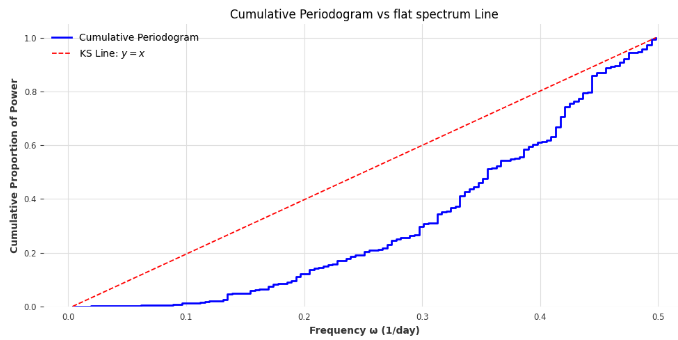


Figure 29: Cumulative Periodogram vs flat spectrum Line

4.6.6 Interpretation/Is this brownian noise

To statistically test for white noise, the Kolmogorov–Smirnov (KS) test can be used. This test compares the empirical cumulative distribution of power to the expected uniform distribution. In this case, the maximum deviation from the uniform line was calculated to be around 76.85% (as shown in the printed output of the code), which is greater than the typical threshold of 10%. Thus, the null hypothesis of white noise is rejected.

Conclusion: Based on both the visual deviation in the cumulative periodogram and the result of the KS-type test, the time series does not resemble white noise. This confirms that further modeling is appropriate.

```
Maximum deviation from uniform: 76.85%
Conclusion: Not white noise
```

Figure 30: KS result

5 Time Series Interpretation

The goal of all this work is to analyze the weekly energy consumption data and understand its behavior.

Stationarity

To check if the data was stable over time, I used two tests: KPSS and ADF. KPSS said the data is already stationary, so I didn't apply differencing and kept $d = 0$ for ARIMA models. I did try differencing but it didn't make a difference.

ACF and Cumulative Periodogram

The ACF plots helped identify the structure of the series. ACF slowly faded away after lag 1, which suggested that an AR model should be used and that MA is not needed. An AR(1) model could work well (but maybe not for long term forecasting). I tested AR models from order 1 to 5. AIC and BIC scores were lowest for AR(1), meaning it was the best among those. This was also the case for MAPE scores.

The cumulative periodogram showed that my time series is not white noise.

Seasonality

STL decomposition didn't show clear seasonality. But the periodogram showed a strong signal every 52 weeks. So seasonal models like SARIMA with $s = 52$ make sense.

From my current undertanding, the series is stationary and has a repeating pattern every 52 weeks. AR(1) is a solid baseline, while an ARMA(5,2) might be better for longer term forecasting, and a SARIMAX should be the best model when including external data like temperature. I'll be testing them soon.

6 Models

6.1 Baseline Model

The baseline model employed is the *Naïve Drift model.* Unlike the standard naïve model (Yt+1 = Yt), which assumes no change from the last observed value, the drift model linearly extrapolates future values using the trend observed in the training window. Specifically, it forecasts the value at time $T + h$ as:

$$\hat{y}_{T+h} = y_T + h \cdot \frac{y_T - y_1}{T - 1}$$

where y_T is the last observed value in the training set, y_1 is the first, T is the length of the training window, and h is the forecast horizon. This model serves as a simple benchmark for evaluating the performance of more sophisticated forecasting methods.

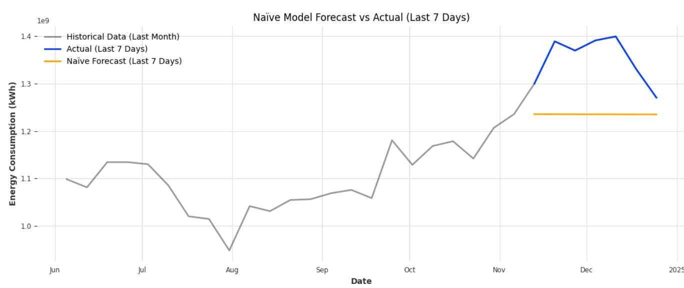


Figure 31: Baseline Naive model, 1 week daily forecast

6.2 AR

6.2.1 Definition

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

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

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

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

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

6.2.2 Plot

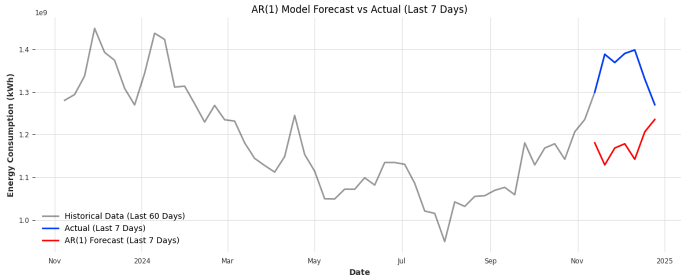


Figure 32: AR(1,0,0) model, 1 week daily forecast

6.2.3 Plot

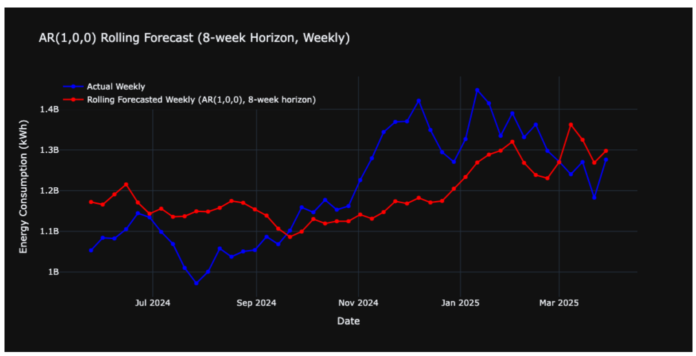


Figure 33: AR(1,0,0) model, 8 week daily forecast

6.2.4 Likelihood ratio test

6.2.5 Model comparison

According to Davison's book [4], a model $f_A(y)$ is nested within a model $f_B(y)$ if B may be reduced to A by restricting certain of the parameters.

- for example, a model $Y_1, \ldots, Y_n \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma^2)$ is nested within the model that the observations are from a Gaussian AR(1) process, because the first is obtained from the second by setting $\alpha = 0$.

Obviously the maximised log likelihoods satisfy $\ell_B \geq \ell_A$, because the more comprehensive model B contains the simpler model A.

The likelihood ratio statistic for comparing A with B is

$$W = 2(\ell_B - \ell_A).$$

If the model is regular, the simpler model is true, and B has q more parameters than A, then

$$W \dot{\sim} \chi^2_q.$$

```
Likelihood Ratio Test (LRT) Results (comparing AR(p) vs AR(p-1)):
AR(2) vs AR(1):  LR stat = 51.173, p-value = 0.0000
AR(3) vs AR(2):  LR stat = 37.705, p-value = 0.0000
AR(4) vs AR(3):  LR stat = 37.527, p-value = 0.0000
```

Figure 34: Likelihood-Ratio Test results for AR

6.2.6 Residuals

Standardized residuals are defined as

$$\tilde{e}_t = \frac{(y_t - \mu_t) - \alpha(y_{t-1} - \mu_{t-1})}{\sigma}, \quad t = 2, \ldots, n,$$

where $\mu_t = \mu + \delta I(t > 38)$.

The residuals should be approximately (Gaussian) white noise.

In the next slide we plot some diagnostics based on $\tilde{e}_2, \ldots, \tilde{e}_n$:

- original data with fitted mean,

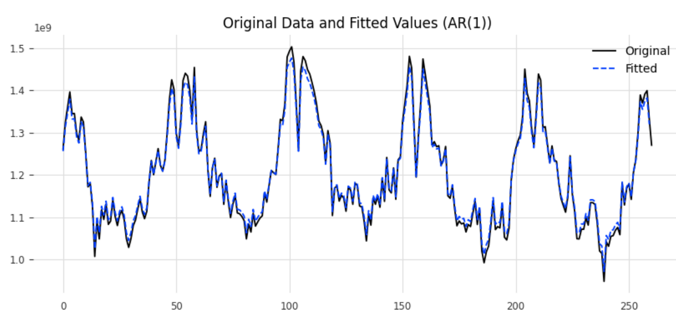


Figure 35: original data with fitted mean, AR model

- time series of residuals,

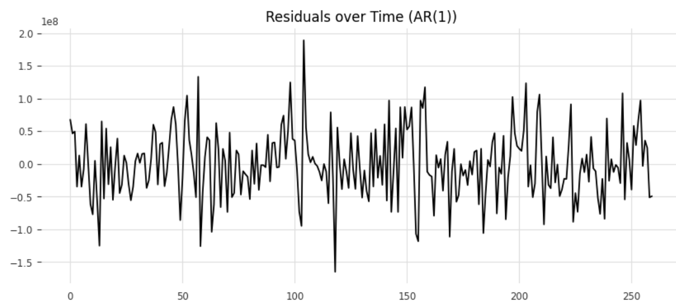


Figure 36: Residuals over Time(AR1)

6.3 ARMA

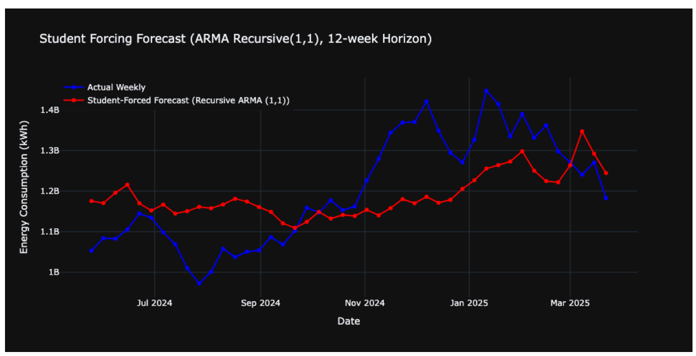


Figure 37: Recursive ARMA over 8-week horizon

6.3.1 ACF & PACF

According to the Time Series Analyis book, To summarise: for causal and invertible ARMA models the ACF and PACF have the following properties:

	AR(p)	**MA(q)**	**ARMA(p,q)**
ACF	Tails off	Cuts off after lag q	Tails off
PACF	Cuts off after lag p	Tails off	Tails off

This gives an approach to identifying AR and MA models based on the ACF and PACF, and suggests how to choose p or q. [4]

6.3.2 ARIMA

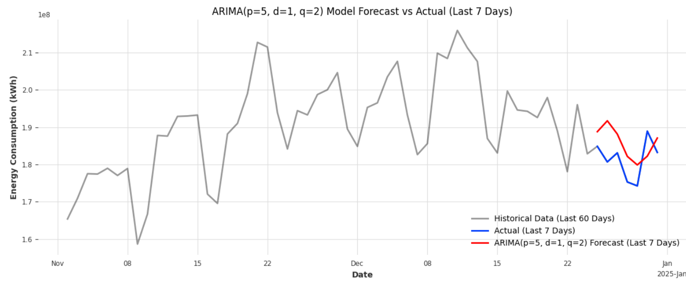


Figure 38: ARIMA 5 first attempt

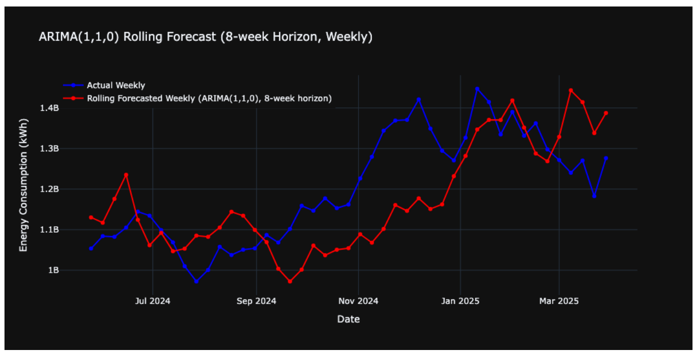


Figure 39: ARIMA(1,1,1) Rolling Forecast

Choice of d:

- Examine a plot of the data for non-stationarity.
- If the data seem non-stationary, we difference successively until they appear stationary.
- Usually $d = 1, 2$ is enough!

Choice of p and q:

- Examine the ACF and PACF of the differenced data.
- Sharp cut-off in the ACF after q lags suggests using an MA(q) model.
- Sharp cut-off in the PACF after p lags suggests using an AR(p) model.
- No sharp cut-off suggests ARMA model, hopefully with $p, q \leq 2$.
- No (or very slow) decline of ACF/PACF to zero suggests need to difference further, or to think again.

– If in doubt, opt for a parsimonious model, with fewer parameters. [4]

In this case, I have chosen ARMA(1,1) for now with d = 0.

6.4 SARIMA

6.4.1 Definition

Many geophysical time series have seasonal components. For example,

– hourly temperatures have 24-hour and annual cycles,

– monthly temperatures have a 12-month cycle.

For seasonal components the period is fixed and known (unlike cyclic behaviour). It may be useful to use an s-fold difference operator $I - B^s$, for example with $s = 12$ to remove the seasonal component from monthly temperatures.

The multiplicative seasonal autoregressive moving average model SARIMA$(p, d, q) \times (P, D, Q)_s$ is

$$\Phi_P(B^s)\phi(B)(I - B)^d(I - B^s)^D Y_t = \alpha + \Theta_Q(B^s)\theta(B)\varepsilon_t,$$

where $\{\varepsilon_t\}$ is Gaussian white noise. The ordinary autoregressive and moving average components are represented by the operators $\phi(B)$ and $\theta(B)$, respectively; the seasonal autoregressive and moving average components by $\Phi_P(B^s)$ and $\Theta_Q(B^s)$, of orders P and Q; and the ordinary and seasonal difference components by $(I - B)^d$ and $(I - B^s)^D$ of orders d and D.

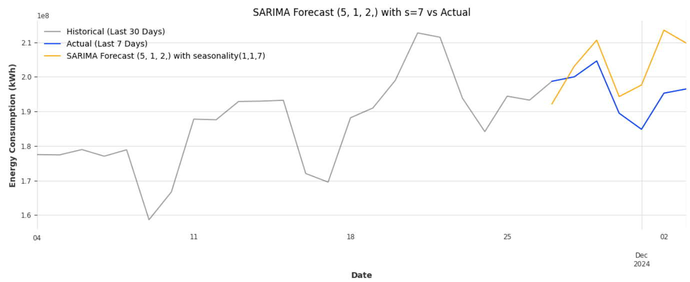


Figure 40: SARIMA 5 first attempt

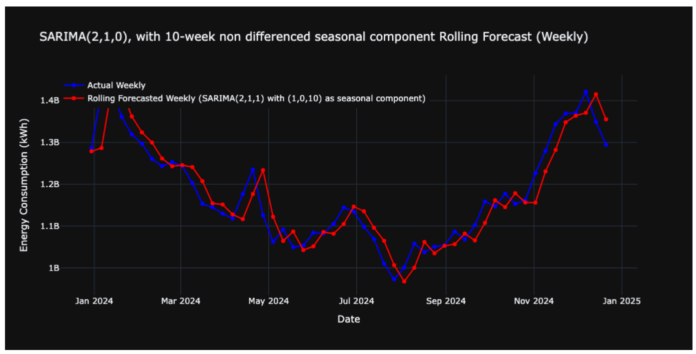


Figure 41: SARIMA(2,1,0), with 10-week non differenced seasonal component Rolling Forecast

6.4.2 Modeling procedure

6.4.3 Residuals

For residuals, I've decided to plot the cumulative periodogram and calculate their KS value, to determine if they are white noise.

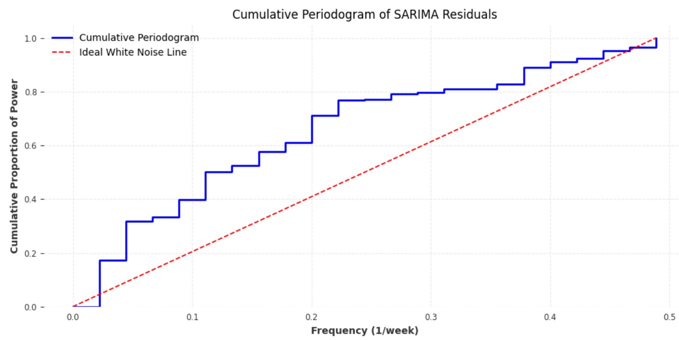
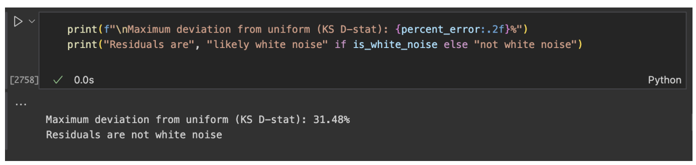


Figure 42: Cumulative Periodogram of SARIMA residuals



```
print(f"\nMaximum deviation from uniform (KS D-stat): {percent_error:.2f}%")
print("Residuals are", "likely white noise" if is_white_noise else "not white noise")
```

[2758] ✓ 0.0s Python

```
Maximum deviation from uniform (KS D-stat): 31.48%
Residuals are not white noise
```

Figure 43: KS Results

The KS results show that the residuals are not white noise, meaning more work needs to be done in the future.

6.5 SARIMAX

6.5.1 Definition and Exogenous Variables

A SARIMAX model is a linear model that generalizes ARIMA by adding support for exogenous variables (covariates), which help improve forecasting by accounting for influences external to the target series.

The general formula of a SARIMAX model is:

$$y_t = \phi(B)^{-1}\theta(B)\Phi(B_s)^{-1}\Theta(B_s)\left((1-B)^d(1-B^s)^D y_t - X_t\beta\right) + \epsilon_t$$

Where:

- y_t: target time series
- B: backshift operator
- d, D: non-seasonal and seasonal differencing orders
- $\phi(B), \Phi(B_s)$: non-seasonal and seasonal AR polynomials

– $\theta(B), \Theta(B_s)$: non-seasonal and seasonal MA polynomials
– X_t: matrix of exogenous variables
– β: vector of coefficients for exogenous variables
– ϵ_t: white noise error term

6.5.2 Weather Data

To enrich the SARIMAX model, we have incorporated weather data from MeteoSwiss. Specifically, the average daily temperature across the Swiss federation as an exogenous variable.

```
Fetching: https://data.geo.admin.ch/ch.meteoschweiz.klima/nbcn-tageswerte/nbcn-daily_SMA_previous.csv
            date  value    city
0     1864-01-01   -0.8  Zurich
1     1864-01-02   -8.0  Zurich
2     1864-01-03  -11.1  Zurich
3     1864-01-04  -10.2  Zurich
4     1864-01-05  -11.3  Zurich
...          ...    ...     ...
58800 2024-12-27   -1.8  Zurich
58801 2024-12-28   -1.5  Zurich
58802 2024-12-29   -1.7  Zurich
58803 2024-12-30   -2.3  Zurich
58804 2024-12-31   -2.5  Zurich

[58805 rows x 3 columns]
Fetching: https://data.geo.admin.ch/ch.meteoschweiz.klima/nbcn-tageswerte/nbcn-daily_SMA_current.csv
          date  value    city
0   2025-01-01    1.1  Zurich
```

Figure 44: Imported Weather Data from MeteoSwiss

6.6 Result

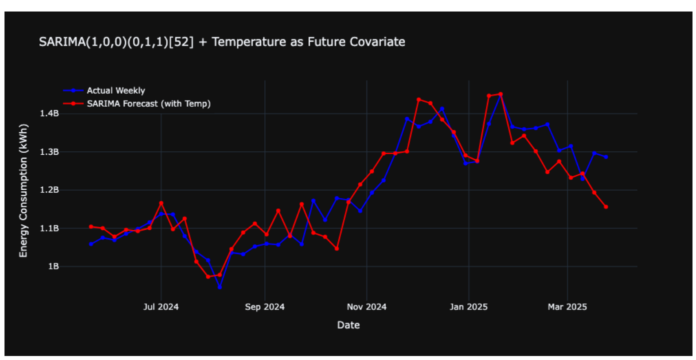


Figure 45: 8 week forecast by SARIMAX

7 Evaluation

7.1 MAPE

7.1.1 Definition

The mean absolute percentage error (MAPE) is a metric used to evaluate the accuracy of a forecasting model. It calculates the average of the absolute percentage errors between the predicted values and the actual values. Lower MAPE values indicate more accurate forecasts.

The formula is:

$$\text{MAPE} = \frac{100\%}{n} \sum_{t=1}^{n} \left| \frac{y_t - \hat{y}_t}{y_t} \right|$$

where y_t is the true value and $\hat{y}_t$ is the predicted value at time t. [10]

7.1.2 Results

Model Comparison – MAPE Scores

Model	MAPE (%)
AR(1)	8.01
ARMA(1,1)	8.12
ARIMA(1,1,1)	8.16
SARIMA(1,0,0 [0,1,1,52])	3.53
SARIMAX(1,0,0 [0,1,1,52]) with Temp	3.86

Figure 46: MAPE comparison between models

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

[10] Wikipedia contributors. Mean absolute percentage error — wikipedia, the free encyclopedia, 2024. Accessed: 2025-06-02.

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