Sardinia Energy Demand

Giovanni Zecchini Jacopo Cesari Tommaso Botticelli

1 Introduction

1.1 Project Statement and Objective

In today's interconnected world, regional energy systems are critical infrastructures whose stability underpins societal and institutional functioning. Anticipating energy demand fluctuations is essential to ensure service continuity and reduce vulnerabilities. As part of stress testing by the government cyber security unit, this project explores scenarios where intentional disruptions could occur without triggering public alarm or systemic failure.

We aim to identify the one-hour time window among the following three days, with the least visible impact on regional activity in the event of a blackout. Sardinia was selected as the target area due to its isolated and manageable energy ecosystem. The analysis focuses on the upcoming days—from 00:00 on February 28th to 23:59 on March 2nd—in preparation for a potential stress test to be conducted during this window.

1.2 Data Description

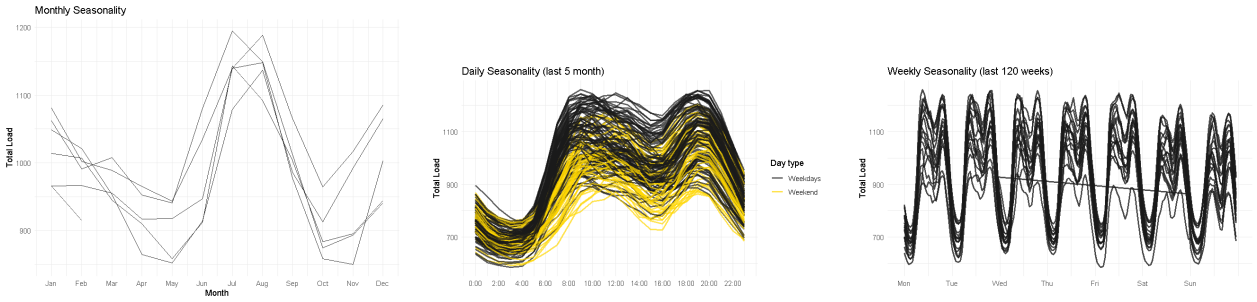
The dataset used in this analysis was obtained from the Terna Download Center, the official data hub of the Italian Transmission System Operator, which provides detailed information on Italy's electricity system. Specifically, we extracted hourly electricity demand data for the Sardinia region covering the period from January 1, 2020, to March 2 2025, yielding approximately 45,000 observations.

In addition to the electricity demand data, temperature for Sardinia over the same period was sourced from the Italian Air Force Meteorological Service. These two datasets were then merged to create a comprehensive dataset integrating both energy demand and temperature variables.

2 Exploratory Data Analysis

2.1 Temporal Aggregations (hourly, daily, weekly, monthly)

As part of the EDA, after observing that our variable of interest was approximately normally distributed, our initial objective was to visually explore the temporal patterns present in the electricity demand data. Given the characteristics of energy consumption, it was expected that distinct seasonal patterns and cycles would emerge on different time scales.



(a) Monthly Seasonality (b) Weekly Seasonality (c) Daily Seasonality

Figure 1: Multi-scale seasonal patterns

To investigate this, we aggregated the original hourly data into three levels: hourly, daily, and monthly averages. In addition, we examined weekly patterns by restructuring daily data into weeks to capture medium-term seasonality.

These temporal aggregations allowed us to generate a series of graphs (Fig. 1), illustrating demand fluctuations over time at multiple granularities.

Clear seasonal patterns are evident at both monthly and hourly scales. Additionally, a consistent decrease in energy demand is observed during weekends.

2.2 Autocorrelation Analysis (ACF and PACF)

Next, we analyzed the temporal dependencies in the data by computing the autocorrelation function (ACF) and partial autocorrelation function (PACF). These diagnostic tools allow for the visual assessment of potential trend and seasonal patterns, informing the appropriate model specification.

The ACF (Fig. 2) exhibits a slowly decaying pattern with regular fluctuations, which is indicative of strong seasonality, while the PACF highlights significant lags that inform subsequent modeling decisions.

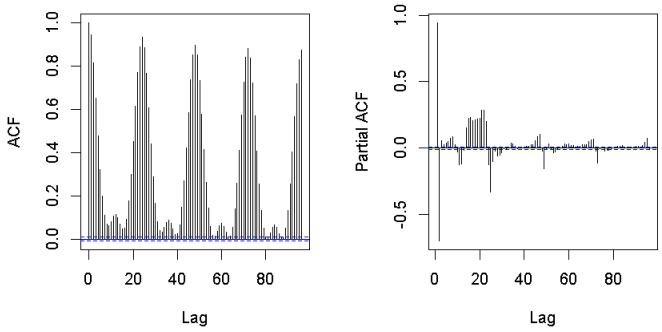


Figure 2: Energy demand time seris ACF and PACF

We conducted stationarity tests to check for stochastic patterns in the data. The Augmented Dickey-Fuller test with various specifications (no intercept, intercept only, intercept, and trend) consistently showed no presence of unit roots, indicating that the time series does not exhibit stochastic trends. Furthermore, the CH test confirmed that seasonal variation is deterministic and stable, with no evidence of stochastic patterns.

2.3 Trend Extraction and Decomposition

To better understand the components of the series, we applied the STL decomposition. In addition, we used moving averages for 4 and 12 months (Fig.3), to smooth the data and highlight long-term trends.

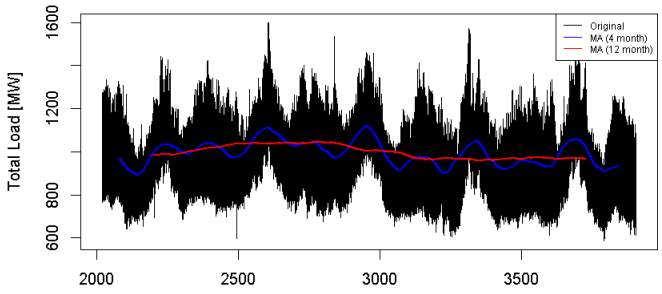


Figure 3: Energy demand time series with moving averages

We decided not to model the structural break. Based on the decomposition plot, we tested for breaks within the interval **Dec 31, 2021 – Jan 31, 2022** by estimating regressions with $D_t = \mathbf{1}\{t > \tau\}$ and storing the F-statistic, where the null hypothesis is $H_0 : \boldsymbol{\beta}' = 0$. We were

unable to identify any break within the interval that aligned with expectations, nor did we find economic, political, or other factors to support the existence of a structural change.

3 Forecasting Models

3.1 Methodology

To conduct a comprehensive and robust analysis of the prediction of energy consumption in Sardinia, we adopted a systematic approach exploring different modeling paradigms, from simple and intuitive methods to more structured and complex approaches.

The objective was to build a comprehensive comparative framework that enabled identification of the 2-3 best performing models through multiple evaluation criteria. The selection of optimal models is based on the following:

- Information criteria: AIC and BIC to balance goodness of fit and model complexity
- Predictive accuracy: Mean Squared Error calculated on rolling validation windows
- Residual analysis: normal distribution, autocorrelation and zero mean check
- Graphical analysis: visual comparison of forecasts with observed values

The dataset was split into training, validation, and test sets. Model evaluation on the validation set was based on MSE, complementing AIC and BIC in the selection process. The adopted validation methodology utilizes rolling window forecasting with a window size of 1.5 years and 72-hour forecasting horizons, iterated over a 15-day validation period. The test set was kept untouched for the final comparison between the top two models.

3.2 Simple Algorithms

Simple models represent the starting point of the analysis and serve as benchmarks to evaluate the effectiveness of more sophisticated approaches. Three baseline models were implemented: *Seasonal Naive*, *Mean*, and *Conditional Mean* (based on seasonality). Among these, the *Seasonal Naive* with a seasonal period of one year (8759 hours) defined as $y_t = y_{t-8759}$ performs best in terms of Mean Squared Error.

However, we chose not to proceed with it because, being too simple, it is overly sensitive to short-term changes, which can distort its forecasts and thus reduce its reliability.

3.3 Exponential Smoothing (ETS)

The ETS model analysis was performed on a time series with a frequency of 24 to capture daily seasonality. Six model configurations were tested using the `ets()` function in R that enables the reformulation of a recursive algorithm into state-space models with different combinations of trend and seasonality components: Simple Exponential Smoothing (ANN), Holt Damped (AAN), ANA, AAA, MAM, and MNM.

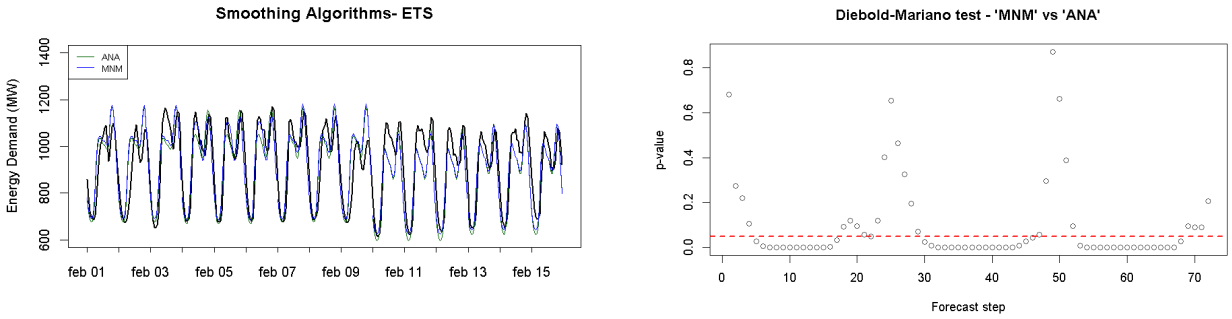


Figure 4: Comparison ets 'MNM' and 'ANA' point forecast

The 'MNM' model shows the lowest AIC and BIC values (783562;783797), suggesting the best balance between fit and parsimony. While, in our view, the 'ANA' model logically better fits the distribution of the time series data and is expected to more accurately capture the underlying dynamics of the phenomenon. The Diebold-Mariano test removes doubts by showing that the MNM model provides superior forecasting performance, at least for most steps ahead.

3.4 Linear Regression Models

3.4.1 Base model

Linear regression model construction follows an incremental approach to identifying the most appropriate specifications. Ten base models were tested with different combinations of regressors. The first regression model we built incorporates a set of core regressors that capture, in our opinion, the essential deterministic components of electricity demand:

$$\text{Total load}_t = \beta_0 + \beta_1 T_t + \beta_2 \text{Hour}_t + \beta_3 \text{Weekday}_t + \beta_4 \text{Month}_t + \beta_5 \text{Festivity}_t + \beta_6 \text{Temperature}_t + \varepsilon_t \quad (1)$$

Here, T_t denotes a linear time trend ($t = 1, 2, \ldots, N$); Hour_t, Weekday_t, and Month_t are categorical variables indicating hour of day (0–23), day of week (1–7), and month of year

(1–12), respectively. Festivity_t is a binary indicator equal to 1 on Italian national holidays, and Temperature_t is the average temperature in Sardinia at time t.

Starting from this initial set of regressors, we fitted different linear models by progressively incorporating nonlinear transformations of the trend (quadratic and logarithmic), interaction terms (such as Hour × Weekday and Hour × Festivity), and trigonometric components (sin/cos) to capture smooth seasonal dynamics. Each specification was evaluated using AIC and BIC to ensure a balance between goodness of fit and model complexity. The model that achieved the lowest values for both criteria is:

$$\begin{aligned}\text{Total Load}_t = {} & \beta_1 \cdot \text{Trend}_t + \beta_2 \cdot \text{Trend}_t^2 + \beta_3 \cdot \log(\text{Trend}_t) + \beta_4 \cdot \text{Hour}_t + \beta_5 \cdot \text{Weekday}_t \\ & + \beta_6 \cdot \text{Festivity}_t + \beta_7 \cdot (\text{Hour} \times \text{Festivity})_t + \beta_8 \cdot (\text{Hour} \times \text{Weekday})_t \\ & + \beta_9 \cdot \text{Month}_t + \beta_{10} \cdot (\text{Weekday} \times \text{Festivity})_t + \beta_{11} \cdot \text{Temperature}_t + \varepsilon_t\end{aligned} \tag{2}$$

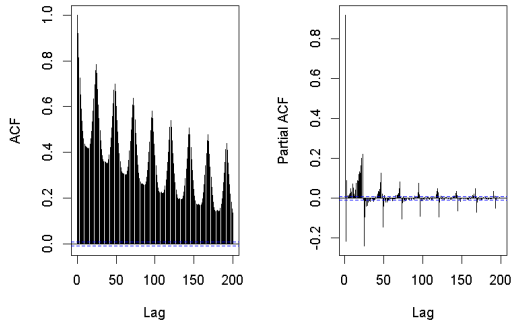


Figure 5: ACF and PACF of linear model residuals

However, the ACF and PACF plots revealed strong autocorrelation and clear seasonal patterns in the residuals, particularly at regular lags, suggesting that relevant temporal dependencies remain unaddressed. To explicitly capture these structures and improve forecast accuracy, we modeled the residuals of the linear regression using various ARIMA and SARIMA specifications.

3.4.2 Dynamic regression

As introduced in the previous chapter, to reduce computational burden we adopted a two-step dynamic regression approach instead of One-Stage Estimation, which estimates all model parameters simultaneously through a single maximum likelihood procedure. In the first step, we estimated a linear model including main regressors. In the second, we modeled residuals using various ARIMA and SARIMA specifications, using as benchmark residuals of the Base model in Eq. 1. For both the complex (Eq. 2) and simple (Eq. 1) versions,

SARIMA(3,0,1)(1,0,2)[24] specification emerged as most suitable, according to Box Jenkins approach, capturing seasonal structure (period = 24) and yielding significantly lower AIC compared to ARIMA alternatives. Among the two, however, the complex model consistently outperformed the simple one in terms of information criteria and forecasting accuracy (MSE). The Diebold-Mariano test (Fig.6) confirms this superiority is statistically significant for short- to medium-term forecasts, though the advantage diminishes at longer horizons, confirming its status as the preferred specification for the remainder of the analysis.

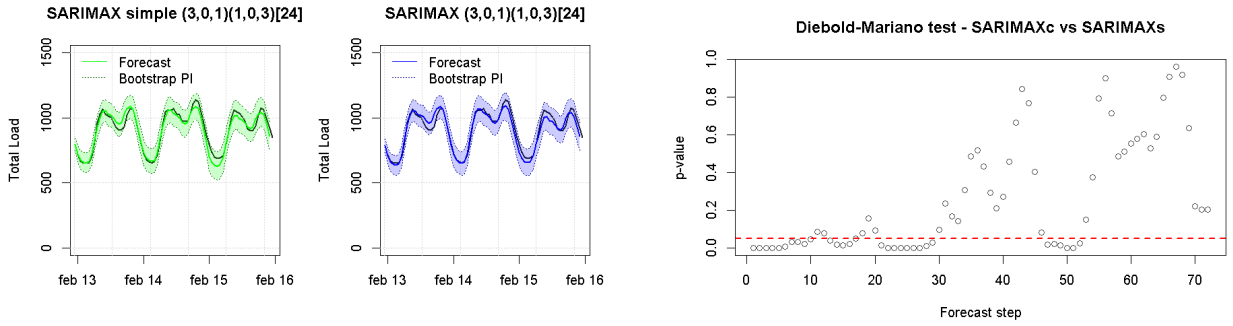


Figure 6: Point forecast comparison between SARIMAX s and SARIMAX c

Residual analysis (Fig. 7) suggests the SARIMAX model captures autocorrelation structures, as evidenced by the ACF and PACF plots. Forecasts appear unbiased, with residuals centered at zero. However, the final plot indicates deviation from residual normality.

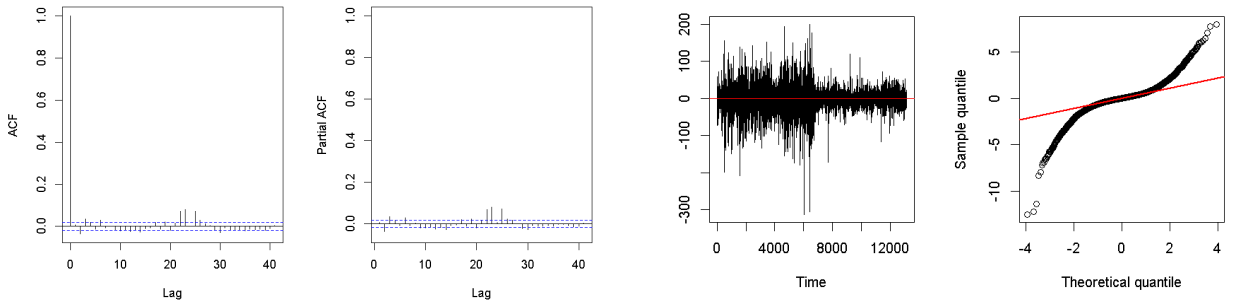


Figure 7: Analysis of residuals SARIMAX c (3,0,1)(1,0,2)[24]

4 Pseudo OOS Model evaluation

Based on the comprehensive analysis conducted, two emerge as final candidates:

1. ETS MNM: optimal smoothing algorithm according to information criteria
2. LM + SARIMAX(3,0,1)(1,0,2)[24]: best model among the SARIMAX configurations

4.1 Point and interval forecast

This section critically evaluates the two best-performing models using a 9-day holdout test set to ensure objective comparison. A 1.5-year rolling window, advanced by 216 hours, was used with a 72-step forecasting horizon. To ensure a realistic evaluation of the model, for the temperature covariate, we used predicted values as input. Forecast accuracy was measured for both point estimates and prediction intervals. Point forecast performance was measured via the MSE $= \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$, where y_i and $\hat{y}_i$ denote observed and predicted counts, respectively. To test the significance of predictive differences, we applied the DM test under the null hypothesis of equal performance. Prediction intervals were evaluated via unconditional coverage, verifying whether empirical coverage approximated nominal $\alpha = 0.95$. Correct coverage is a minimal requirement for interval reliability.

To reduce computational costs, parameters of the LM and SARIMAX(3,0,1)(1,0,2)[24] (dynamic regression) were estimated once on the entire training set (train+validation) outside the iterative loop. In contrast, the ETS MNM algorithm was re-estimated at each iteration to capture potential regime shifts. Within the loop, point forecasts were produced via the `forecast()` function in R.

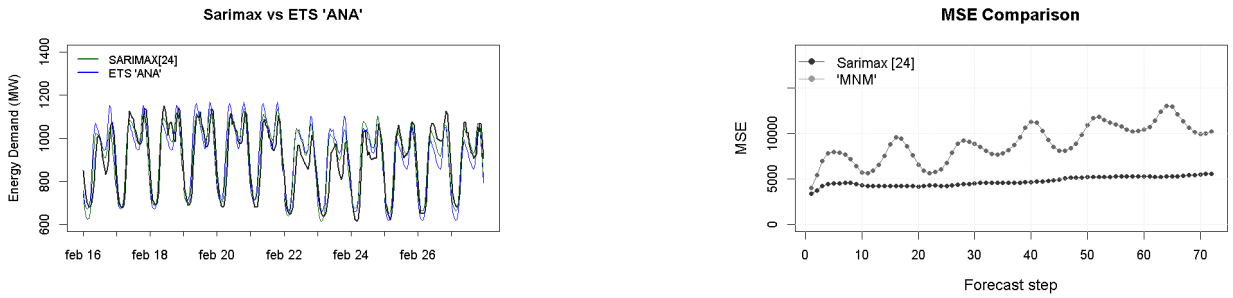


Figure 8: Evaluation of point forecast, comparison of MSEs

In (Fig. 8), we report the point forecasts of SARIMAX and ETS 'MNM' for comparison. To avoid overcrowding the plot, forecasts were displayed every 72 iterations for the 72-steps-ahead horizon of interest. To better illustrate model performance, the right plot in Fig. 8 presents temporal dynamics of the Mean Squared Error for the SARIMAX and ETS `MNM` models, highlighting the consistent superiority of SARIMAX.

ETS prediction intervals were directly obtained from the `$upper` and `$lower` components. Two-stage SARIMAX prediction intervals were computed by combining the variance of the regression component with the forecast error variance; intervals are thus derived from this aggregated uncertainty. Furthermore, the evolution of p-values from the DM test applied to the MSE confirms the outperformance of SARIMAX at each forecasting step. The Diebold-

Mariano test (`dm.test`) was employed to formally compare predictive accuracy of the two models. The alternative hypothesis was specified as `alternative = "greater"`, testing whether the MNM model exhibits a significantly higher mean squared error (i.e., worse predictive performance) compared to SARIMAX.

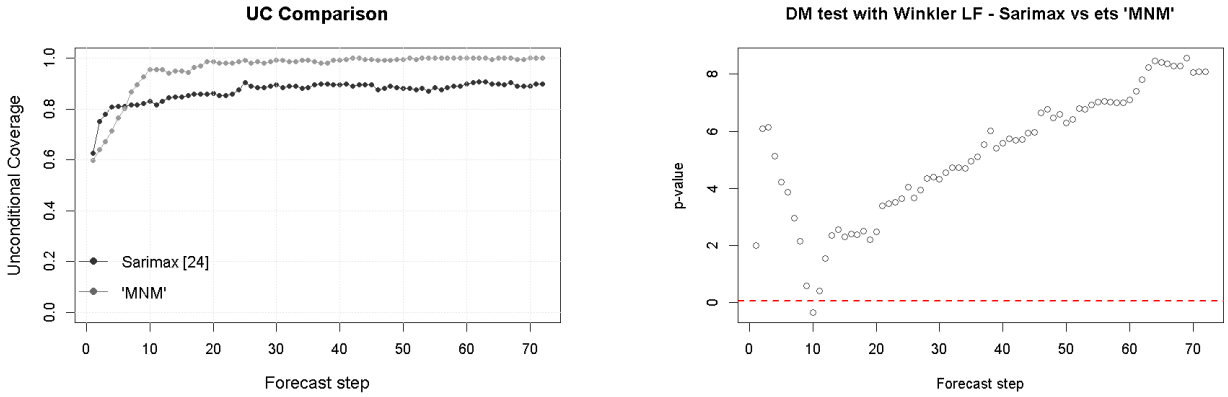


Figure 9: Evaluation of prediction intervals

Regarding prediction intervals, Fig. 9 shows unconditional coverage. For SARIMAX, UC stabilizes around 87–90% instead of 95% target. This undercoverage was expected since prediction intervals were constructed under the assumption of normality, an approximation adopted due to large volume of data available and to reduce computational costs. In contrast, ETS `MNM` shows the opposite issue: UC fluctuates between 95–100%, reflecting underestimation of uncertainty for short horizons and overestimation for longer ones.

Winkler Loss function provides a comprehensive metric balancing interval width and reliability. A lower Winkler LF indicates intervals that are narrow yet adequately cover the true values. The Winkler score function demonstrates that SARIMAX intervals remain more stable across forecast horizons. This is confirmed by the following graph in Fig. 9, which displays the dynamics of the DM test p-values comparing the Winkler scores of the two models (SARIMAX and ets 'MNM'), confirming the previous evidence that SARIMAX achieves significantly better interval performance.

5 Temperature forecast

To include temperature as an explanatory variable in the electricity demand forecasting models, it was necessary to predict hourly temperatures during both the model evaluation phase and the actual forecasting period. Meteorological data were obtained from the Italian Air Force (Aeronautica Militare), based on observations from five weather stations distributed

across Sardinia. The regional hourly temperature was computed as the arithmetic mean of the measurements from these stations. Forecasts of hourly temperatures were then produced and used as inputs in the demand models for both performance evaluation and final forecasting.

The EDA revealed strong annual and daily seasonality in the temperature series. The daily seasonality is confirmed by the ACF, which shows recurring oscillations every 24 lags, while the annual seasonality is shown by the STL decomposition. The data distribution is approximately normal. Moreover, the ADF test rejects the presence of a unit root, excluding a stochastic trend, while the CH test suggests a constant deterministic seasonal pattern. For this reason, we fitted several linear models to the data and selected the one with the lower AIC. The selected model specification incorporates linear and quadratic trend terms, hourly and monthly fixed effects, as well as interaction effects between hour and month, and between hour and trend. We then modeled the residuals using various SARIMAX specifications. Based on ACF, PACF, AIC, and BIC, the best-performing model was a SARIMAX(1,0,3)(2,0,0)[24]. Finally, we compared the MSE of this model with two simpler benchmark, seasonal mean and seasonal naive, using an expanding window approach. The seasonal mean outperformed the others (Fig.10), and was therefore used to forecast temperatures.

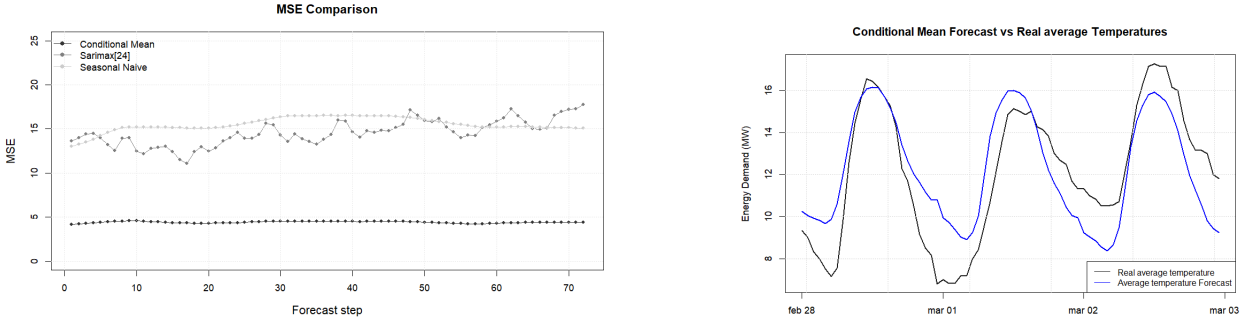


Figure 10: Point forecast evaluation

6 Conclusions

Using SARIMAX(3,0,1)(1,0,2)[24], we forecasted 72 hours ahead to identify the period with minimum expected energy consumption for optimal blackout test timing. The analysis identified `2025-02-28 03:00:00` as the predicted minimum, instead of `2025-03-02 04:00:00`. Although the forecast does not exactly match the observed minimum within the three-day horizon, the difference in predicted energy demand is only 49 MW. Most of this error appears to result from the model's failure to correctly identify the weekend days, that would require more advanced modeling techniques, not examined in this course.