PGR304 – Predictive Analytics

Candidate Number 37

1 Understanding the dataset

The dataset is an hourly multivariate time series of the power system. The dataset covers the period from 31.12.2014 to 31.12.2018 (with 35 064 time stamps). Each row corresponds to one hour in the power system, and all five variables are measured at that same time:

- `Load_MW`: Total electricity consumption (our target).
- `Temp_C`: Ambient temperature, driver for heating and cooling.
- `Price_DA_EUR_MWh`: Day-ahead market price, an economic signal of supply and demand.
- `Solar_MW`: Solar production, affects net load, zero at night.
- `Clouds_pct`: Cloud coverage in percentage, affects solar and temperature.

1.1 Summary statistics

The summary statistics show that:

- `Load_MW`: has an average around 28 696.93 MW, min 18 041 MW, max 41 015 MW and median around 28 901 MW.
- `Temp_C`: mean around 17.63, min around -4.31, max 38.0 and median around 17.04.
- `Price_DA_EUR_MWh`: mean around 49.87 £/MWh, min 2.06, max 101.99 and median around 50.52.
- `Solar_MW`: mean around 1432 MW, min 0, max 5792 MW, median around 616 MW.
- `Clouds_pct`: mean around 20.74%, min 0%, max 100% and median around 20%.

8 rows × 5 columns pd.DataFrame

	load_MW	temp_C	price_DA_EUR_MWh	solar_MW	clouds_pct
count	35028.000000	35064.000000	35064.000000	35046.000000	35064.000000
mean	28696.939905	17.633954	49.874341	1432.665925	20.740475
std	4574.987950	7.234569	14.618900	1680.119887	25.604016
min	18041.000000	-4.319344	2.060000	0.000000	0.000000
25%	24807.750000	12.000000	41.490000	71.000000	0.000000
50%	28901.000000	17.040000	50.520000	616.000000	20.000000
75%	32192.000000	23.000000	60.530000	2578.000000	20.000000
max	41015.000000	38.000000	101.990000	5792.000000	100.000000

Figure 1: Descriptive statistics for the five variables in the dataset.

```
load_MW: min=18041.0, max=41015.0, mean=28696.94, median=28901.00
temp_C: min=-4.319343749999973, max=38.0, mean=17.63, median=17.04
price_DA_EUR_MWh: min=2.06, max=101.99, mean=49.87, median=50.52
solar_MW: min=0.0, max=5792.0, mean=1432.67, median=616.00
clouds_pct: min=0, max=100, mean=20.74, median=20.00
```

Figure 2: Printed summary of minimum, maximum, mean and median for each variable.

1.2 Scale differences and possible outliers

One thing that we can see from the boxplots is:

- `Load_MW` in the 10 000s.
- `Solar_MW` in the 100s-1000s.
- `Price_DA_EUR_MWh` around the 20s up to the 100s.
- `Temp_C` around -4 to 40.
- `Clouds_pct` from 0-100%.

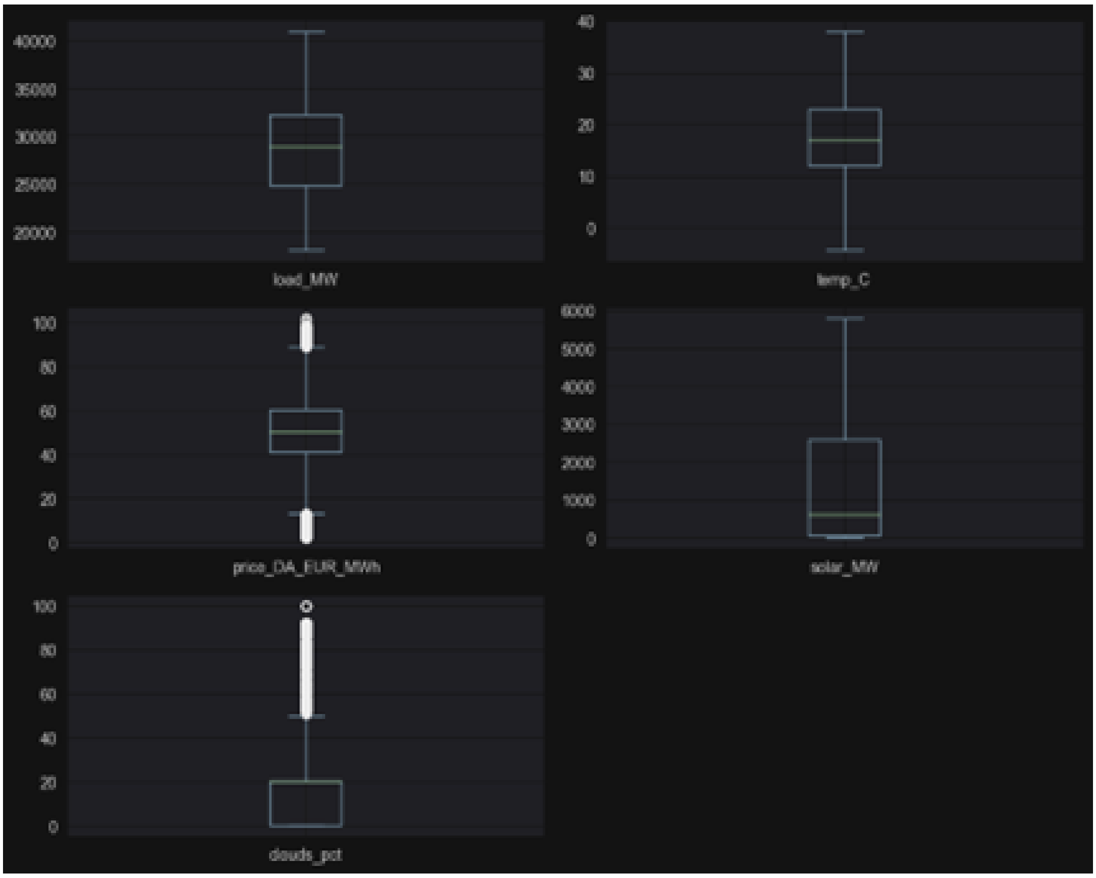


Figure 3: Boxplots illustrating scale differences and possible outliers for all variables.

1.3 Visual exploration

After loading and cleaning the missing values, exploration was conducted on the time series. This helps uncover patterns that are not obvious from the raw hourly data tables. In the notebook, there are multiple visualizations that aggregate the hourly data into daily and monthly averages. These plots reduce hourly noise, making it easier to see long-term patterns that are harder to detect in hourly data.

1.4 Time plot of daily mean load vs temperature

The plot displays the daily load against the daily temperature. Daily mean temperature shows a seasonal trend, decreasing in winter and increasing in summer. We can see that the daily load decreases slightly as temperature increases, although there are still some peaks. In general, there is more load in colder periods compared to warmer periods.

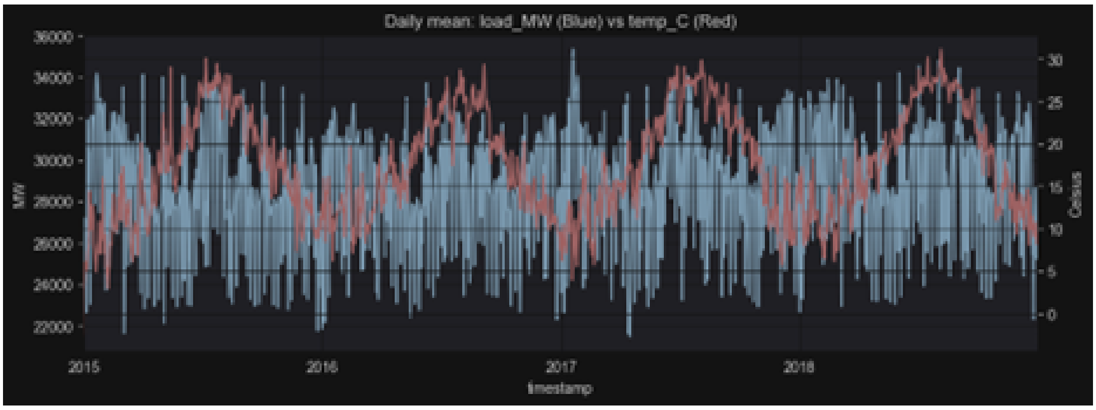


Figure 4: Time plot of daily mean `Load_MW` (blue) versus `Temp_C` (red).

1.5 Daily mean load vs price in EUR

The plot displays load versus price per MW. We can see that high load often correlates with high prices, but prices also have some independent spikes.

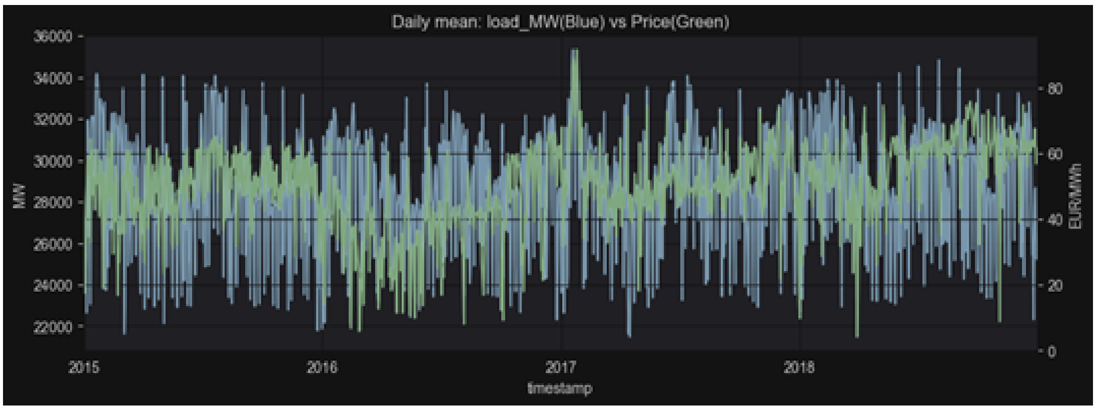


Figure 5: Time plot of daily mean `Load_MW` (blue) versus `Price_DA_EUR_MWh` (green).

1.6 Monthly means

Monthly exploration smooths the series even more and makes seasonal patterns easier to see. Monthly temperature clearly repeats a winter low and summer high. This helps justify including seasonal terms in forecasting models. In addition, we can see that the solar power follows almost the same pattern as temperature, as there is more solar power during summer than during winter.

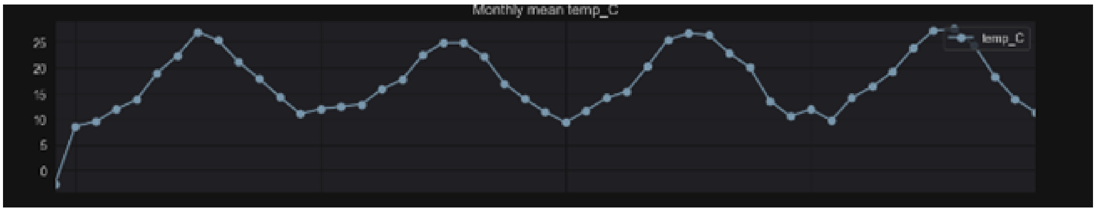


Figure 6: Monthly mean temperature (`Temp_C`).

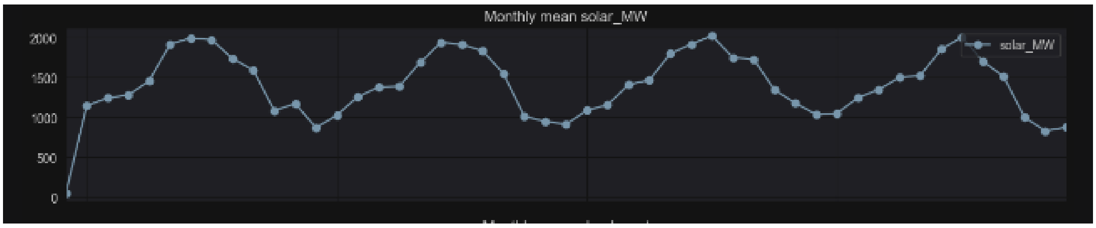


Figure 7: Monthly mean solar production (`Solar_MW`).

1.7 Monthly mean load vs temperature

Monthly mean load versus temperature shows that there is some kind of inverse relationship: when temperature rises, load tends to fall. The red line is temperature, which is high in summer months, while the load at the same time is for the most part lower, as it is warmer with longer sun duration. There are some unexplainable factors during the summer, as there are still some months with high load even though it is not needed as much; this could be affected by rainy days or cloudy days during the warmer periods.

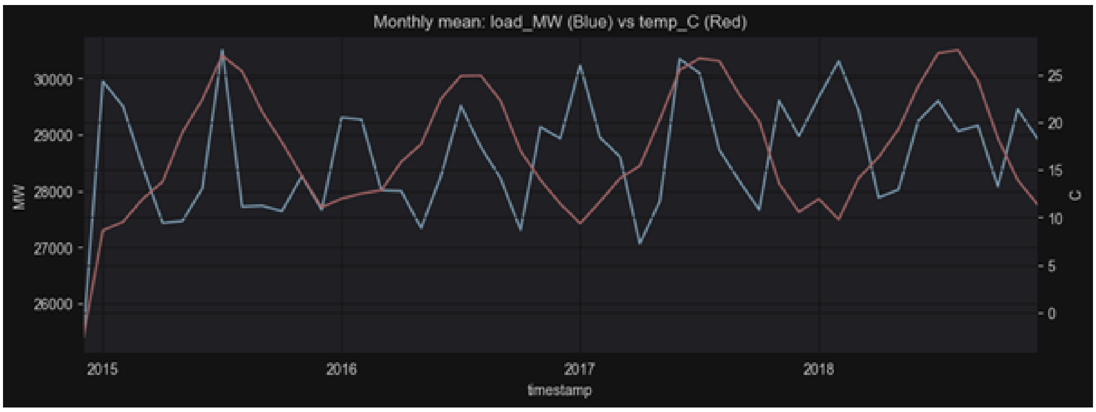


Figure 8: Monthly mean `Load_MW` (blue) versus `Temp_C` (red).

1.8 Monthly mean load vs price

Monthly mean load vs price shows that price moves with load over longer periods, where months with higher load generally have higher prices, even though price also reflects market conditions and supply and demand.

The daily and monthly visualizations illustrate a clearer picture of seasonality and relationships between variables than the raw hourly plots. They show that load, temperature and solar

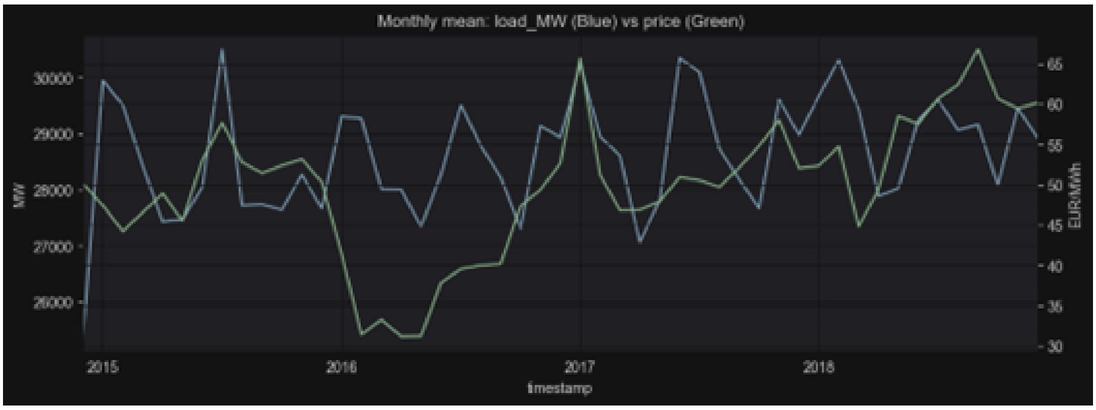


Figure 9: Monthly mean `Load_MW` (blue) versus price (green).

follow seasonal cycles, and that load is to some degree negatively correlated with temperature and solar generation. Prices tend to move with load but are also influenced by other factors. Clouds remain random, even at the monthly level, indicating that they have limited predictive power on their own.

2 Tools and methods used

The analysis was carried out in Python using a Jupyter notebook. The main libraries were `pandas` for data loading, cleaning and transformation, `NumPy` for numerical operations, and `matplotlib` for all figures, including descriptive plots, boxplots and time–series visualisations. The time-series models and statistical tests were implemented with the `statsmodels` package.

Methodologically, the work follows a standard forecasting workflow. First, basic descriptive statistics and visualisations were produced to understand the level, spread and seasonal patterns in all five variables. This includes summary tables, boxplots, time plots of daily and monthly averages, and pairwise correlation matrices.

Second, temporal dependence and stationarity were examined. Autocorrelation functions (ACF) were used to reveal the strong 24-hour cycle. Formal stationarity tests (Augmented Dickey–Fuller and KPSS) were then applied before and after 24-hour seasonal differencing. Granger-causality tests and cross-correlation functions (CCF) were used to investigate whether temperature, price, solar production and cloud cover contain predictive information for future load.

Third, two forecasting model families were estimated: multivariate Vector Autoregression (VAR) models on the seasonally differenced multivariate series and univariate seasonal ARIMA (SARIMA) models for the load series alone. Lag-order selection and model comparison relied on information criteria such as AIC and BIC.

Finally, forecast accuracy was evaluated on a 48-hour test set using mean absolute error (MAE) and mean absolute percentage error (MAPE). These metrics provide both an error measured in MW and a scale-free percentage measure, and they are used consistently to compare the VAR and SARIMA forecasts.

3 Utility value

This section will examine how the explanatory variables relate to the target variable `Load_MW`, and whether the data should be modelled as a multivariate model (VAR) or a univariate model (ARIMA, SARIMA). The analysis is based on (1) correlation analysis, (2) Granger causality and (3) cross-correlation function (CCF).

3.1 Correlation analysis

The correlation matrix shows pairwise correlations between load and the four predictors. Figure 10 displays the full correlation matrix, and Figure 11 shows the correlations with `Load_MW` sorted by value.

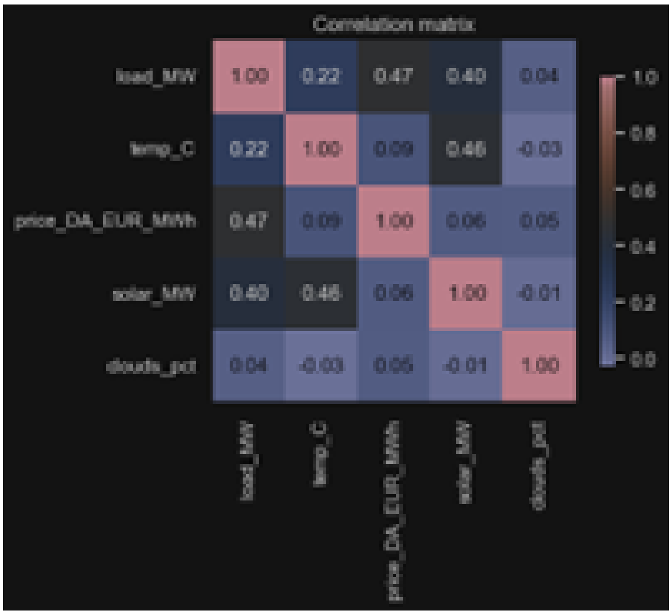


Figure 10: Correlation matrix between `Load_MW` and the four explanatory variables.

```
Correlation with load_MW (descending):
price_DA_EUR_MWh    0.474273
solar_MW            0.395511
temp_C              0.220790
clouds_pct          0.040361
```

Figure 11: Correlations with `Load_MW` sorted in descending order.

From this analysis:

- The day-ahead price has the highest correlation with load (approximately 0.47). This is reasonable, as prices tend to increase when demand is high.
- Solar generation has a relatively high correlation with load (approximately 0.40), mainly due to common seasonal patterns in power systems.
- Temperature is also positively correlated with load (approximately 0.22). Even though lower temperatures generally increase electricity demand, the long hourly series contains many overlapping patterns, so the pure temperature–load relationship appears weaker than for price and solar.
- Cloud cover is almost uncorrelated with load (approximately 0.04), so it is not expected to be a strong standalone predictor.

From pure correlation, we can see that the most informative predictors are price and solar. Temperature is useful but weaker, and clouds contribute only a small amount.

3.2 Correlation among predictors

When looking at the four predictors:

- We can see that temperature and solar production have some correlation between them ($\approx$ 0.46). This explains the shared weather/season signal (warmer and sunnier periods often occur around the same time).
- Day-ahead price has weak correlation with the weather variables ($\approx$ 0.06-0.09), which means price adds some information that is not already in temperature or solar.
- Clouds coverage is close to zero with all of them ($\approx$ 0.03-0.05).

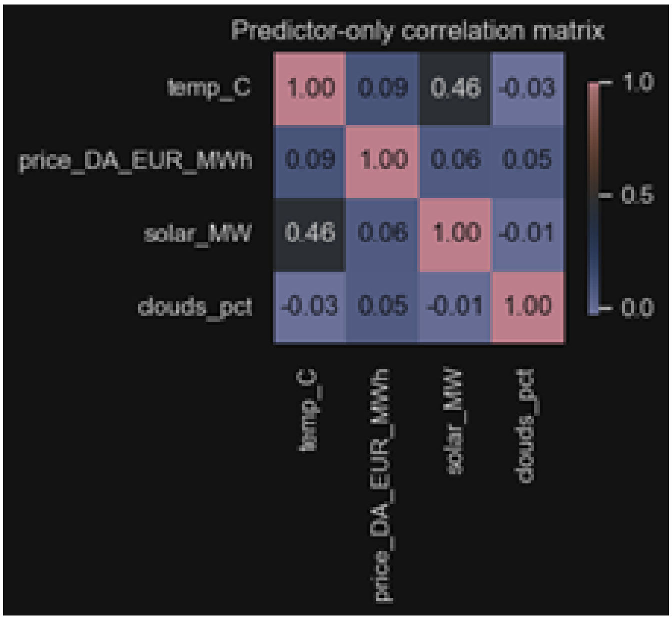


Figure 12: Correlation matrix among the four predictors.

This matters for later modelling, as all predictors can be included in one model, but we should be aware that temperature and solar partly describe the same underlying patterns.

3.3 Granger causality test

To assess the model selection, a pairwise Granger causality test was applied to all variables using a 24-hour lag. The reason behind the 24-hour window is that the data are hourly, and load usually reacts to weather and price information within the same day, so one full daily cycle is enough to capture relevant lead–lag effects while keeping the model compact. The resulting heat map shows each predictor–response pair, using the lowest p-value found across those 24 lags.

Since the dataset is very large (35 000 hourly points) and all series share strong daily and seasonal patterns, many variables show up as statistically significant at the 5% level. In the Granger test, temperature, day-ahead price and solar production show significance within 24 hours. Cloud coverage appears to be the only variable that does not show any statistical significance.

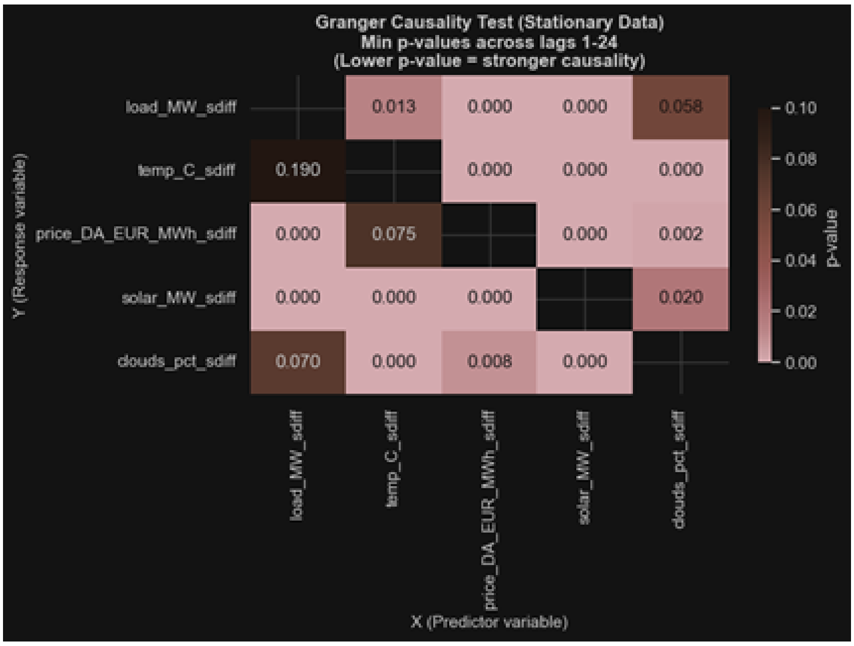


Figure 13: Granger causality p-values over 24 lags for all predictor-response pairs (lower p-values indicate stronger Granger causality).

The most important point is that there is predictive information in several variables for future load. This supports using a multivariate VAR model with most of the predictors included. In particular, price, temperature and solar show significant predictive information for future load. However, cloud cover does not show significant contribution and can be excluded from the model to simplify it without losing accuracy.

3.4 Cross-Correlation Function (CCF)

To compare and complement the Granger-causality tests, a cross-correlation function (CCF) analysis was conducted on the seasonally differenced data. The CCF measures how two series move together at different time lags. Positive CCF values indicate that the predictor and load move in the same direction, while negative values mean they move in opposite directions. Peaks at negative lags show that the predictor leads the response, whereas peaks at positive lags indicate the response leads the predictor. Since the CCF computes correlation at each lag, it is symmetric and does not establish causality.

The CCF plots reveal that the day-ahead price has a strong positive correlation with load at lag 0, indicating that price and load move together. Temperature, solar production and cloud cover show weaker correlations, with their peaks showing at negative lags around -13 to -20 hours. These negative values reflect that increases in temperature or solar production are associated with decreases in load. The interpretation of the CCF described above—the sign of the values indicating whether two series move together or in opposite directions, and the sign of the lag indicating which series leads—is consistent with guidance on cross-correlation functions for time-lagged series published by Esri's ArcGIS documentation[5].

Comparing CCF and Granger results shows that variables with significant Granger p-values (price, temperature and solar) also show peaks in the CCF plots, although the correlations for temperature and solar are small. This is expected, as Granger causality tests are stricter and can detect predictive relationships even when correlation is weak. A variable can Granger-cause another with either a positive or negative effect, so weak or negative CCF values do not rule out predictive power. Cloud cover has a marginally significant Granger p-value and a very small CCF peak, suggesting it may contribute less to forecasting than the other variables.

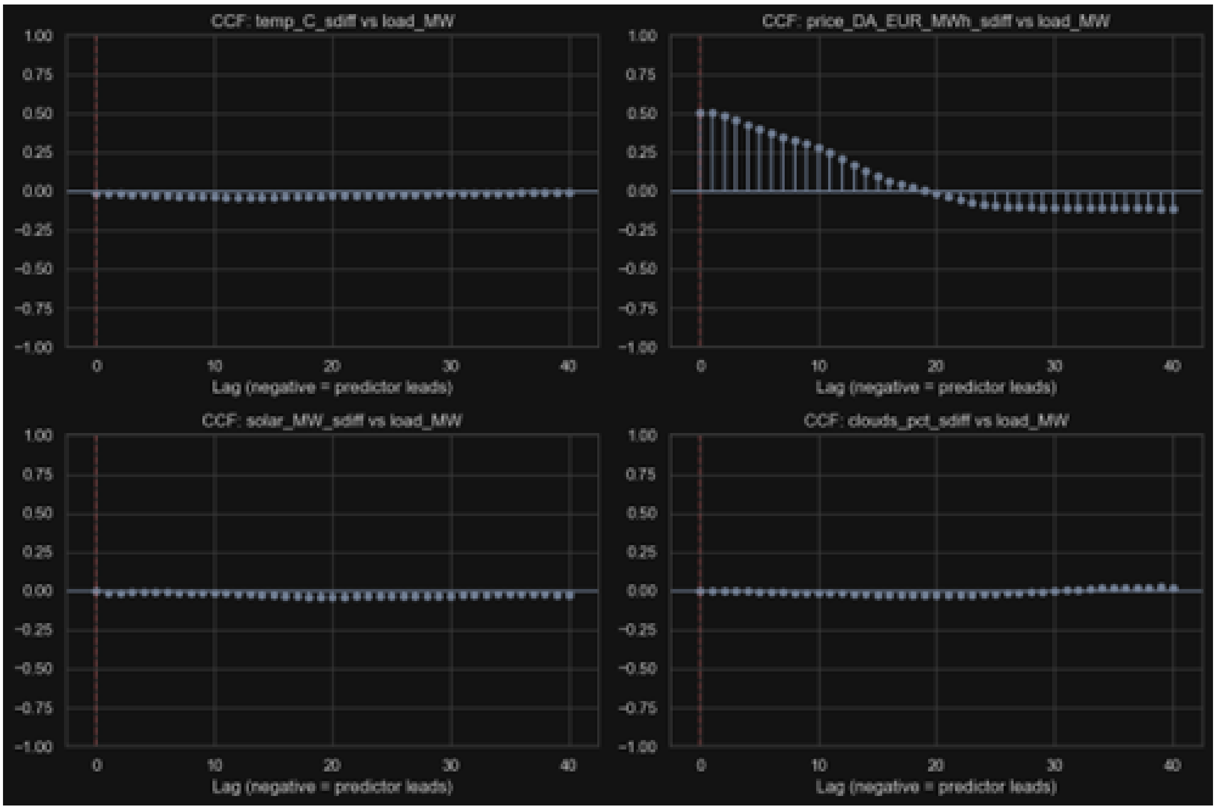


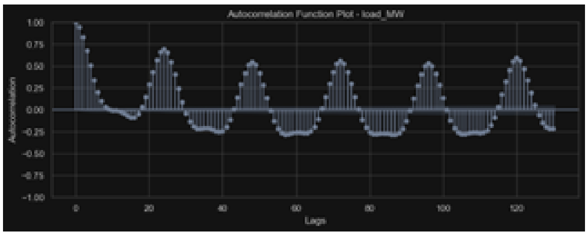
Figure 14: Cross-correlation functions between seasonally differenced predictors and `Load_MW`.

4 Preprocessing and stationarity

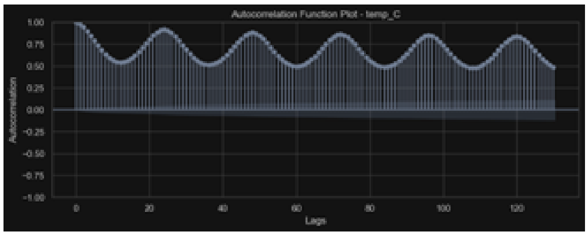
This section explores whether the dataset should be used directly for forecasting, and if not, which preprocessing steps are necessary. The dataset consists of five time-aligned hourly variables. Time series models like VAR assume that the input series are stationary. The analysis below shows that the raw hourly series is not fully stationary and must be transformed.

4.1 Data preprocessing, ADF, KPSS and ACF

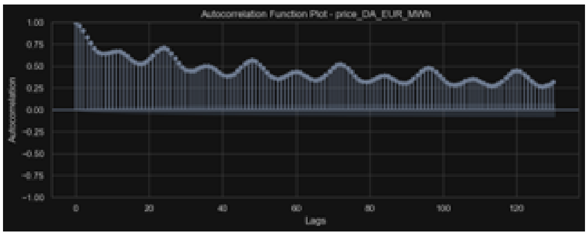
The autocorrelation function (ACF) plots for all variables show repeating spikes at every 24 lags, especially for load and solar production. These patterns are for hourly data points and indicate a 24-hour seasonal component. The values at a given hour are highly related to the value at the same hour on the previous day. This indicates that the series cannot be considered stationary for time series modelling and that transformations are needed before fitting time series models such as VAR or SARIMA.



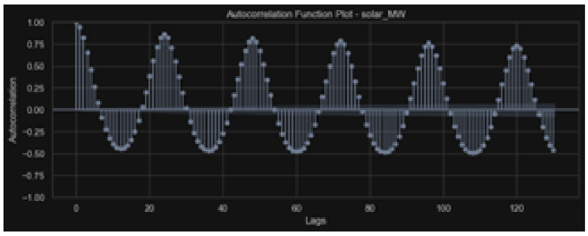
(a) ACF of `Load_MW`



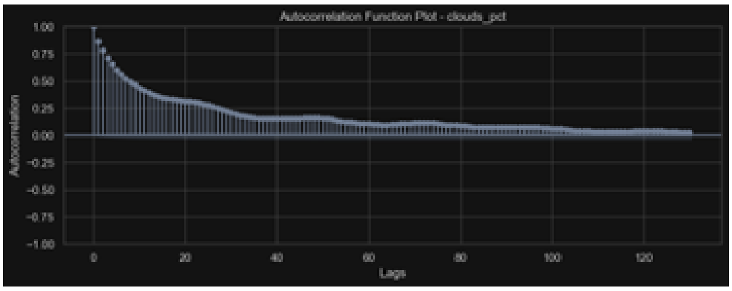
(b) ACF of `Temp_C`



(c) ACF of `Price_DA_EUR_MWh`



(d) ACF of `Solar_MW`



(e) ACF of `Clouds_pct`

Figure 15: Autocorrelation function (ACF) plots for all five variables.

4.2 ADF and KPSS

In addition to the ACF, two other stationarity tests were conducted:

- **Augmented Dickey-Fuller (ADF)**: For all variables, the ADF test statistic was significant, suggesting that the dataset is stationary.
- **Kwiatkowski-Phillips-Schmidt-Shin (KPSS)**: For all variables except cloud cover, the KPSS test indicates that the data are not stationary.

It is common that ADF and KPSS disagree like this. The two tests have opposite null hypotheses: the ADF test assumes a unit root under the null, while the KPSS test assumes stationarity under the null. Statsmodels' stationarity and detrending guide outlines how these tests can yield different conclusions and how trend-stationary or difference-stationary series should be treated[6]. When they do, and when we also consider the ACF, it is safe to say there is still structure in the data (repeating patterns or trends) that should be removed before modelling. Therefore, the dataset is not yet ready to be used directly in a forecasting model.

4.3 Seasonal differencing

Since the main problem was a repeating daily cycle (24-hour cycle), seasonal differencing with period 24 was applied to all five variables. Seasonal differencing means subtracting today's value from the value at the same hour yesterday[4]:

$$x_t^{(\mathrm{sdiff})} = x_t - x_{t-24}.$$

Even though the stationarity tests indicated that cloud-cover percentage is stationary, seasonal differencing was applied to all variables. The reason for this was to keep the multivariate structure consistent and to make sure all series were transformed in the same way before modelling.

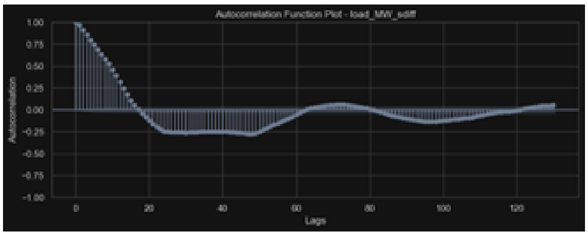
The transformation targets the 24-hour pattern visualised in the ACF and has the advantage that it keeps short-run movements. After differencing, the first 24 rows become missing (because there is nothing 24 hours earlier to subtract), so these rows were dropped. The remaining seasonally differenced series were combined into a new, modelling-ready DataFrame `df_stationary`.

At the same time, the last observed 24 hours of the original, non-differenced data were saved into a separate object. This is necessary for the forecasting stage, where a model predicts values on the differenced scale. The predictions then have to be "added back" to the last observations to return to the original units.

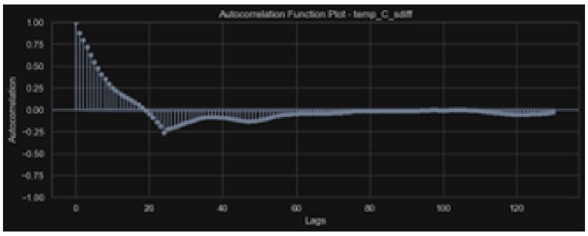
4.4 Re-checking stationarity after transformation

After seasonal differencing, ADF, KPSS and ACF were applied again to the differenced features. The results can be summarised as follows:

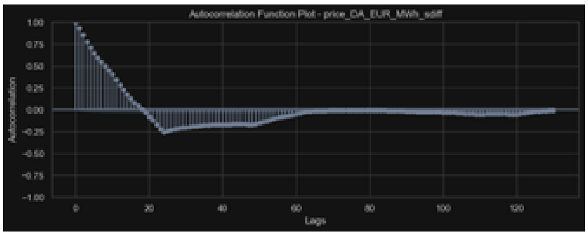
- For all five features, the ADF test rejects the unit root.
- For all five features, the KPSS statistic is small and does not reject stationarity.
- The new ACF plots no longer show large 24-hour waves; instead, the autocorrelation decays and oscillates around zero.



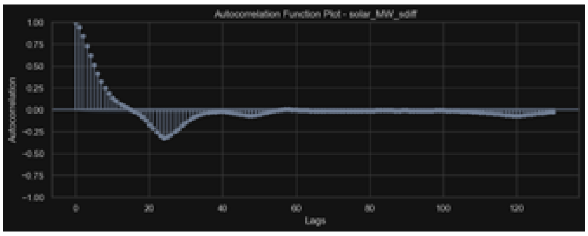
(a) ACF of seasonally differenced `Load_MW`



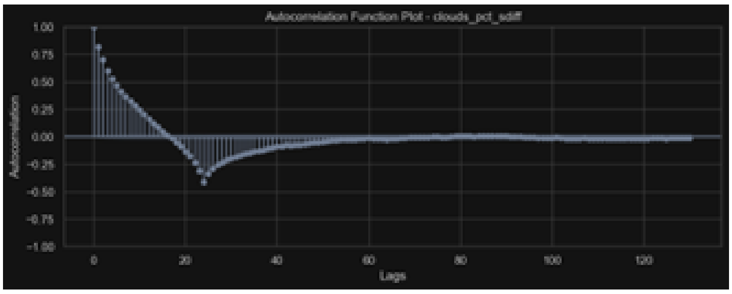
(b) ACF of seasonally differenced `Temp_C`



(c) ACF of seasonally differenced `Price_DA_EUR_MWh`



(d) ACF of seasonally differenced `Solar_MW`



(e) ACF of seasonally differenced `Clouds_pct`

Figure 16: ACF plots after applying a 24-hour seasonal difference to all variables.

The combination of all three diagnostics (ADF, KPSS and ACF) indicates that the data are now stationary, which is what we aim for. The remaining autocorrelation at short lags is normal and is expected to be captured by the lag structure of the final model. After applying the 24-hour seasonal difference, the ACF of all variables still shows a small spike around 24 lags; this is an expected effect of the differencing operation and not a sign of non-stationarity. It simply means the model should include some lags to explain this dependence. Since both ADF and KPSS indicate stationarity and the autocorrelation decays afterwards, no further differencing was applied. In other words, the series are now stationary and ready for modelling.

4.5 Consideration of other transformations

A log transformation was considered to stabilise variance. However, the dataset contains variables that can be zero or negative, such as solar power and temperature at night. A log transform would therefore either fail or require special handling. More importantly, the main issue was seasonality, not variance that grows with the level. Since the seasonal differencing already addressed the main issue and made the series stationary, a log transform was not applied.

5 Analysis, modelling and prediction

In this section the stationary time series is used to build and evaluate forecasting models for load. First, the data are divided into a training set and a test set in a way that respects the time ordering of the observations. Then different model classes (such as VAR) are estimated on the training data and used to produce forecasts. The forecasts are compared on the held out test set so that model performance is assessed on unseen future observations, consistent with how the models would be used in practice.

5.1 Train/test split

After making the stationary dataset using 24-hour seasonal differencing, the dataset was split into training and test sets based on time. For the test set, the last 48 hourly observations of the stationary series are held out, while all earlier observations are used for training. This ensures that the models are always trained on past observations and evaluated on unseen future observations.

- **Train:** stationary but unscaled training data (for models like ARIMA/VAR that can work directly on differenced series).
- **Test:** stationary but unscaled test data for evaluating the models.

5.2 Model selection & justification

All three diagnostics, the correlation matrix, cross-correlation function (CCF) and Granger-causality tests, show that the day-ahead price has the strongest correlation with load, while temperature and solar have weaker (and negative) correlations at lags around 13-20 hours. Even though these effects are small, the Granger tests indicate that temperature and solar still contain predictive information for future load. Cloud cover shows a very weak CCF peak and a non-significant Granger p-value, and will therefore be excluded from the forecasting models.

A univariate AR(p) model (autoregressive model of order p) for the load series y_t (here y_t is `load_MW` at time t) is written as

$$y_t = c + \sum_{i=1}^{p} \phi_i y_{t-i} + \varepsilon_t,$$

where c is a constant, ϕ_i are the autoregressive coefficients, and ε_t is a white-noise error term (a random shock with mean zero and constant variance). An ARIMA(p, d, q) model extends this by applying differencing d times and adding a moving-average part.

For the seasonally differenced series, we define

$$w_t = \nabla_{24} y_t = y_t - y_{t-24},$$

which corresponds to a seasonal difference with lag 24 to remove the daily seasonality. An ARIMA$(p, 1, q)$ model on y_t with this seasonal difference can then be viewed as an ARMA(p, q) model on w_t:

$$w_t = c + \sum_{i=1}^{p} \phi_i w_{t-i} + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} + \varepsilon_t,$$

where the θ_j are the moving-average coefficients. These univariate models exploit the strong autocorrelation and daily pattern in load, but they do not use the additional information from temperature, price and solar production.

On the other hand, a VAR(p) model treats the system as a vector of series

$$\mathbf{y}_t = \left(\text{load\_MW}_t,\ \text{temp\_C}_t,\ \text{price\_DA\_EUR\_MWh}_t,\ \text{solar\_MW}_t,\ \text{clouds\_pct}_t\right)^\top,$$

and models all of them jointly as

$$\mathbf{y}_t = \mathbf{c} + A_1 \mathbf{y}_{t-1} + A_2 \mathbf{y}_{t-2} + \cdots + A_p \mathbf{y}_{t-p} + \boldsymbol{\varepsilon}_t.$$

Here, $\mathbf{c}$ is a vector of constants, each A_i is a 5×5 coefficient matrix that captures how all variables at lag i affect all variables at time t, and ε_t is the vector of white-noise errors.

In the load equation of the VAR, the past values of load, temperature, price and solar production all enter directly. This allows the model to react to changes in these drivers explicitly, instead of only seeing their indirect effect through past load, as in a purely univariate ARIMA model.

5.3 Lag order selection using AIC and BIC

After model selection, the next step is to find the lag order p of the VAR model. A higher lag order allows the model to use a longer history of all variables, but it also increases the number of parameters and the risk of overfitting. To balance goodness of fit and model complexity, the Akaike information criterion (AIC) and the Bayesian information criterion (BIC) were computed on the training data. The VAR model was fitted to the stationary training set for a range of lag orders, and the corresponding AIC/BIC values were compared.

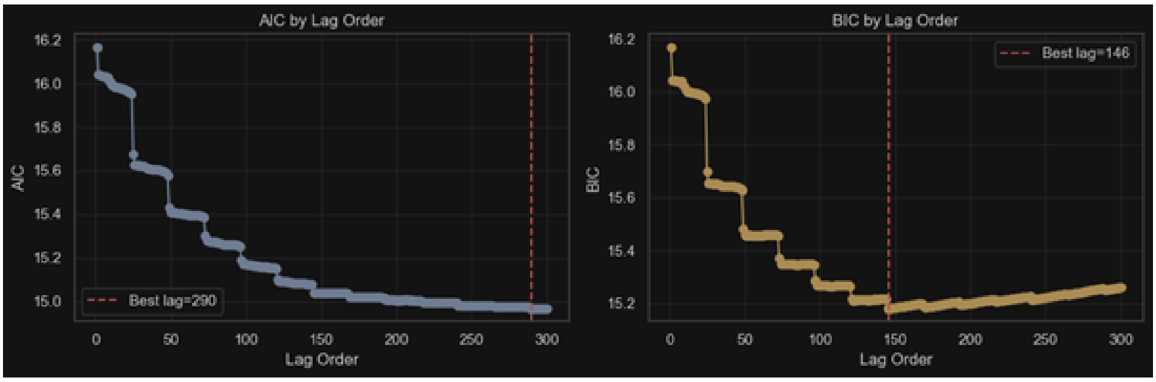


Figure 17: AIC and BIC values by lag order for the VAR model.

5.4 VAR forecasting and model evaluation

After selecting the optimal lag length using AIC and BIC, a VAR model was fitted to the seasonally differenced training data. The AIC and BIC both reached their minimum at lag order 146. For the baseline forecast, the model was trained on all available predictors (load, temperature, day-ahead price, solar generation and cloud cover) and used to generate a 48-hour-ahead forecast. The differenced forecasts were then inverted back to the original scale using the last observed values. The best forecast for the 48 test hours is shown below.

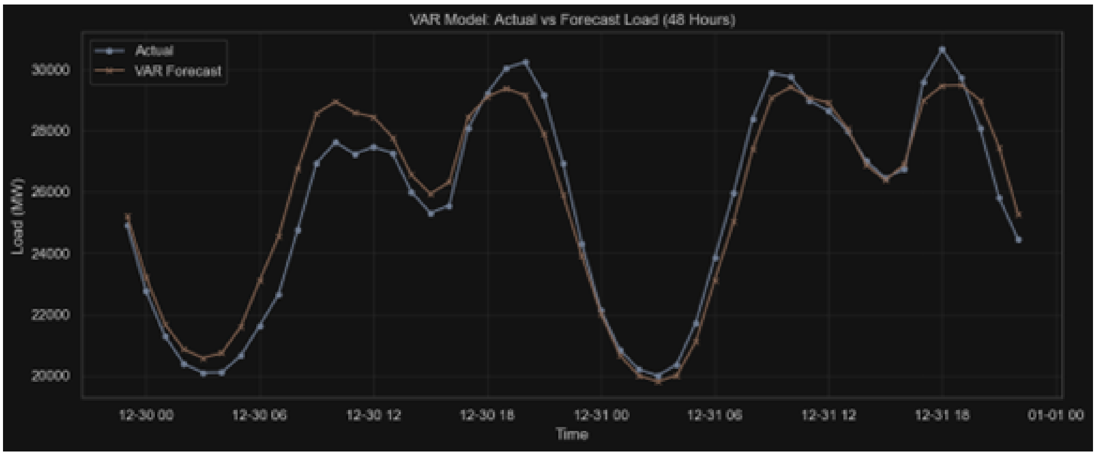


Figure 18: VAR model: actual vs. forecast load (48 hours).

In addition, multiple VAR specifications were estimated for comparison: one excluding cloud cover, one excluding both cloud cover and temperature, and one including only price and load. The error statistics for these models are shown in Figure 19. The model without temperature and cloud cover achieved a lower error than the full model, suggesting that these extra predictors may introduce noise. The load-and-price-only model reduced the complexity even further while still achieving very good results. Overall, the best models were those that included two or three of the variables, typically load together with price and solar.

```
Error Statistics:
  Mean Error: 714.73 MW
  Mean Absolute Error: 1058.47 MW
  Mean Absolute Percentage Error: 4.20%
  Max Overprediction: 3150.40 MW
  Max Underprediction: -1244.15 MW
```

(a) All variables

```
Error Statistics:
  Mean Error: 541.52 MW
  Mean Absolute Error: 1000.81 MW
  Mean Absolute Percentage Error: 3.95%
  Max Overprediction: 3000.96 MW
  Max Underprediction: -1319.99 MW
```

(b) All variables except cloud

```
Error Statistics:
  Mean Error: 167.15 MW
  Mean Absolute Error: 724.91 MW
  Mean Absolute Percentage Error: 2.85%
  Max Overprediction: 1921.39 MW
  Max Underprediction: -1286.61 MW
```

(c) All variables except cloud and temperature

```
Error Statistics:
  Mean Error: 182.86 MW
  Mean Absolute Error: 696.24 MW
  Mean Absolute Percentage Error: 2.74%
  Max Overprediction: 1987.21 MW
  Max Underprediction: -1264.20 MW
```

(d) Price and load only

Figure 19: Error statistics for different VAR specifications.

The analysis demonstrates that multivariate VAR models can capture short-term dynamics in electricity load when combined with appropriate pre-processing and feature selection. Stationarity testing and seasonal differencing were important to remove daily and weekly cycles, while Granger causality and cross-correlation analysis guided the choice of explanatory variables and lag order. Even when forecasting with the full VAR model using all predictors, the predictions were already reasonably good. A more specific and better forecast is obtained by excluding temperature and cloud cover, which yields even lower forecast errors. This suggests that, for this dataset and time period, variation in load is more closely related to price dynamics and past values of load than to weather fluctuations.

In summary, the project presents a complete workflow from data cleaning and exploratory analysis to modelling, forecasting and interpretation. The recommended VAR(146) model excluding all variables except day-ahead price achieved the best results, with mean absolute error (MAE) around 725 MW and mean absolute percentage error (MAPE) of about 2.85%.

6 Alternative approach: SARIMA

A seasonal ARIMA (SARIMA) model is chosen as the univariate choice for load forecasting, as the series shows a strong daily cycle. SARIMA extends ARIMA by including seasonal autoregressive and moving-average terms with a seasonal period of 24 hours. Unlike a VAR model, which uses multiple predictors (price, solar, etc.), a SARIMA model can capture the recurring seasonal pattern using only the load series.

In contrast, a VAR model jointly models the target and predictor series, allowing it to incorporate the relationship between explanatory variables and load. Research suggests that VAR outperforms ARIMA when variables are highly correlated, while ARIMA or SARIMA can match VAR when there are additional variables that provide little extra information [1].

In summary, SARIMA is included to represent the univariate approach: it focuses on the load series alone and exploits its seasonality, while VAR represents the multivariate approach, exploiting relationships between load and the explanatory variables.

6.1 SARIMA model identification

Unlike the manual differencing approach, modern SARIMA estimation can apply both the non-seasonal (d) and seasonal (D) differencing inside the model. There is no need to difference the values before fitting. Instead, the SARIMA order parameters d and D are set to indicate how many differences to take. For example, for hourly data with a 24-hour cycle we set the seasonal period $s = 24$ and usually choose $D = 1$ to remove the daily pattern. The SARIMA implementation in `statsmodels` internally differences the series according to the specified orders and seasonal arguments [3].

Once the differencing orders are fixed, the ACF is inspected to guide the autoregressive (AR) and moving-average (MA) orders. A strong spike at lag 24 in the ACF indicates a seasonal component. Seasonal lags (multiples of 24) in the ACF point to seasonal AR or MA terms, while short lags point to non-seasonal AR or MA terms. Based on this inspection, we consider SARIMA models of the general form

$$\text{SARIMA}(p, d, q) \times (P, D, Q)_{[24]},$$

where (p, d, q) are the non-seasonal orders and (P, D, Q) are the seasonal orders with period 24.

After choosing reasonable candidate orders from the ACF structure, we fit SARIMA models over a grid of (p, d, q, P, D, Q) values and select the best orders using information criteria. Specifically, a loop was used to evaluate different combinations and the model with the lowest AIC or BIC was chosen [4].

By following this approach, we let SARIMA handle all differencing internally and use the ACF as a guide, while information criteria confirm the optimal lag orders. This ensures that the final model captures both short-term and daily seasonal dynamics of the load series.

6.2 Forecast evaluation (MAE and MAPE)

To evaluate forecast accuracy, we use Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE). MAE is the average absolute difference between forecasted and actual values, measured here in MW and reflecting the typical forecast error. MAPE is the average absolute percentage error, giving a scale-free measure of accuracy [2].

For the VAR analysis, the best model (VAR(146) with load and price only) achieved a MAE of about 725 MW and a MAPE of about 2.85% on the 48-hour test set. The SARIMA models produced clearly higher errors: for a representative SARIMA specification the MAE was around 2500 MW and the MAPE around 10%, indicating that the univariate model lacked some predictive information that is present in the explanatory variables.

The summary error statistics for the SARIMA and VAR models are shown below.

```
Error Statistics:
  Mean Error: 2533.79 MW
  Mean Absolute Error: 2533.79 MW
  Mean Absolute Percentage Error: 9.95%
  Max Overprediction: 5146.16 MW
  Max Underprediction: 358.94 MW
```

(a) SARIMA model

```
Error Statistics:
 Mean Error: 182.86 MW
 Mean Absolute Error: 696.24 MW
 Mean Absolute Percentage Error: 2.74%
 Max Overprediction: 1987.21 MW
 Max Underprediction: -1264.20 MW
```

(b) VAR model

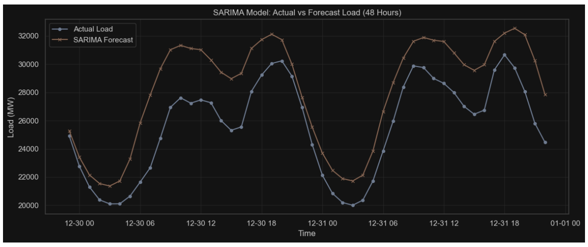
Figure 20: Error statistics (MAE and MAPE) for the SARIMA and VAR models.

Based on these results, the VAR model achieves substantially lower forecast errors than the SARIMA model. This is likely because the VAR uses additional predictors, in particular the day-ahead price, which provide extra predictive information that a purely univariate SARIMA model cannot capture.

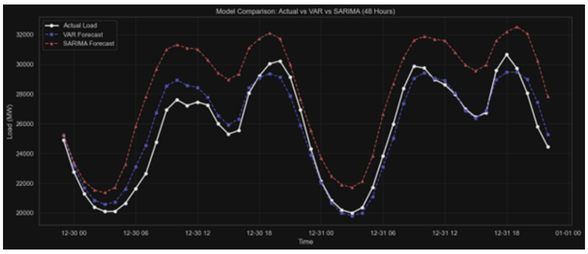
6.3 Forecast comparison

The forecast plots (Figure 21a for SARIMA and Figure 21b for both models) visualise and compare the two approaches. Both forecasts track the overall daily load pattern, but the SARIMA prediction tends to smooth some short-term fluctuations and slightly lags sudden changes. In Figure 21a the SARIMA forecast (orange line) captures broad peaks and shows clear cyclical structure, but occasionally under- or overshoots the actual load during rapid changes.

In contrast, the VAR forecast in Figure 21b more closely follows the actual load line, especially during peak periods. The VAR model's forecasts appear to align better with the peaks and declines because it directly incorporates price signals that correlate with high-demand hours. Overall, visually the VAR forecast tends to hug the true load curve more tightly (less bias around peaks), whereas the SARIMA forecast is smoother. Both models capture the daily seasonality, but the VAR model, through the inclusion of additional explanatory variables, is better able to adjust to short-term deviations from the average pattern.



(a) SARIMA forecast vs. actual load.



(b) VAR and SARIMA forecasts vs. actual load.

Figure 21: Visual comparison of SARIMA and VAR forecasts against actual load.

7 Results and evaluation

This section brings together the main forecasting results and discusses which variables and lags are most influential, as well as the assumptions and limitations that affect how the results should be interpreted.

On the 48-hour test set, the multivariate VAR approach clearly outperforms the univariate SARIMA model. The best VAR specification, a VAR(146) model using load and day-ahead price, achieves a mean absolute error of about 725 MW and a mean absolute percentage error of around 2.85%. In contrast, a representative SARIMA model yields an MAE around 2500 MW and a MAPE of roughly 10%. This means that, on average, the VAR forecasts deviate from the realised hourly load by less than 3%, whereas the SARIMA forecasts are off by about 10% of the true value. The forecast plots show that the VAR model follows peaks and troughs more closely, especially during high-demand hours, while SARIMA tends to smooth sharp changes.

The variable and lag-structure of the VAR models is consistent with the earlier exploratory analysis. Correlation matrices, Granger-causality tests and CCF plots all indicated that price and solar generation are the most informative predictors for load, with temperature weaker and cloud cover almost uninformative. When estimating alternative VAR specifications, the models that excluded cloud cover, and then also temperature, achieved lower errors than the full model. This suggests that, within this dataset, lagged load and lagged price carry most of the predictive signal, while the extra variables mainly add noise. The optimal lag order $p = 146$ found by AIC/BIC is high, reflecting that several days of history improve the fit, but the underlying ACF structure indicates that the most influential lags are the very short lags and multiples of 24 hours corresponding to the daily cycle.

These results rely on several modelling assumptions. First, the analysis assumes that 24-hour seasonal differencing has rendered the series stationary, and that the relationships between load, price and weather are approximately linear and stable over the 2015-2018 period. Second, the VAR framework treats the predictors as given, even though in reality price and load are jointly determined in the power market. Third, model performance is assessed on a single 48-hour test window at the end of the sample. No rolling-origin evaluation or cross-validation was carried out, so the reported errors may not fully represent performance in other periods or under rare conditions such as holidays or extreme weather. Finally, the models do not account for structural changes in technology or demand that might occur outside the observed time span.

Taken together, the forecasts should therefore be viewed as reliable mainly for short-term (one to two day ahead) predictions under conditions similar to those seen in the historical data. Within this range of validity, the VAR(146) model with load and price appears to be a suitable baseline for operational load forecasting in this system. For longer horizons, or in future years where demand patterns or market rules may change, the models would need to be re-estimated and reassessed on fresh data.

8 Conclusion

In this analysis the multivariate VAR model gave better accuracy than the univariate SARIMA model. This suggests that for this dataset, where price and weather signals have some correlation with load exploiting that extra information yields a clear advantage. However, the SARIMA model still performed reasonably well, reflecting that load has strong built-in seasonality that SARIMA can model effectively.

In situations with even stronger and more stable seasonal patterns (and fewer useful predictors), a SARIMA model might perform comparably or even better. In contrast, if more explanatory variables are available, a multivariate approach like VAR would likely outperform a purely univariate SARIMA. In conclusion, VAR worked best for this dataset, but when good covariates are lacking or a simpler model is preferred, SARIMA remains an appropriate choice for capturing clear seasonal dynamics.

References

[1] M. Salauddin and U. Khan, "Comparison of forecasting performance with VAR vs. ARIMA models using economic variables of Bangladesh," *Asian Journal of Probability and Statistics*, vol. 10, no. 2, pp. 33-47, 2020. `https://doi.org/10.9734/ajpas/2020/v10i230243`.

[2] Wikipedia contributors, “Mean absolute percentage error,” *Wikipedia, The Free Encyclopedia*, 12 June 2019. `https://en.wikipedia.org/wiki/Mean_absolute_percentage_error`.

[3] Statsmodels Developers, “`statsmodels.tsa.statespace.sarimax.SARIMAX` - Statsmodels 0.14.0 documentation,” *statsmodels.org*, n.d. `https://www.statsmodels.org/stable/generated/statsmodels.tsa.statespace.sarimax.SARIMAX.html`.

[4] R. J. Hyndman and G. Athanasopoulos, “Seasonal ARIMA models,” in *Forecasting: Principles and Practice*, OTexts, n.d. `https://otexts.com/fpp2/seasonal-arima.html`.

[5] Esri, “How Time Series Cross Correlation works,” *ArcGIS Pro documentation*, n.d. Available online: `https://pro.arcgis.com/en/pro-app/latest/tool-reference/space-time-pattern-mining/how-time-series-cross-correlation-works.htm`.

[6] Statsmodels Developers, “Stationarity and detrending (ADF/KPSS),” *statsmodels 0.15.0 documentation*, n.d. Available online: `https://www.statsmodels.org/dev/examples/notebooks/generated/stationarity_detrending_adf_kpss.html`.