Laboratory 4 – Regression over Autoregressive Forecast

Problem Description

In this laboratory, we aim again to develop accurate predictive models for the current flow in a 5-bus kite network. The network is depicted in Figure 1. The loads generated by wind-power are given and our objective is to predict the currents, this time considering the forecast of the input variables.

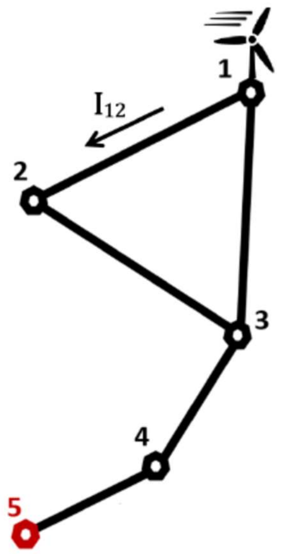

Figure 1: 5-bus kite network

An error is set on the power injections, which are in bus 1, where the wind turbine is depicted. We aim to forecast the current I_{12}.

To address this problem, we employ several predictive modelling techniques, starting with Ordinary Least Squares (OLS) regression. OLS regression allows us to establish a baseline prediction model based solely on wind generation data. However, this initial model may overlook important dependencies and suffer from autocorrelation in the residuals.

Using Durbin-Watson statistics, we tested the dataset for autocorrelation. With a value of 0.197, which is much smaller than 2, it indicates strong positive autocorrelation in the dataset.

To mitigate autocorrelation effects, we apply the Cochrane-Orcutt transformation. This transformation adjusts the regression model to account for autocorrelation in the residuals, resulting in improved prediction accuracy. By iteratively refining the model through the Cochrane-Orcutt procedure, we aim to enhance the robustness of our predictions.

Additionally, we explore the use of autoregression to capture the temporal dependencies inherent in the current flow data. Autoregression allows us to model the current I_{12} as a function of its previous values, incorporating both wind generation and load flow data. By considering the historical trends and patterns in the data, autoregression provides valuable insights into the dynamic behaviour of the current flow in the network.

Furthermore, we enhance the autoregression model by incorporating the sum of all bus load flows. This addition enables us to account for the aggregate effect of load variations on current flow, further improving the accuracy of our predictions. By integrating wind generation, load flow, and their autocorrelation into the predictive modelling framework, we strive to develop comprehensive models that accurately forecast I_{12} in the 5-bus kite network.

Results

In Figure 2 we the results of our different model structures can be compared to the real current value. We can see, that the OLS model is the least accurate, with the highest discrepancy to the real current. This indicates that the OLS model, which solely relies on wind generation data, fails to capture the complex dynamics and dependencies influencing the current flow in the network. The Cochrane-Orcutt transformation we applied on the model, results in a much higher accuracy of the results. This enhancement underscores the effectiveness of mitigating autocorrelation effects through the Cochrane-Orcutt procedure, leading to more reliable predictions of current flow. The autoregressive model using the loads but nor the load sums shows similar results to the OLS model and is only slightly better. However, the error is much higher than with the Cochrane-Orcutt transformation. The relatively higher error suggests that while incorporating load data improves prediction, it may not adequately capture the intricate dynamics of current flow in the network. Finally, the autoregressive model, that also uses the load sums provided the best results by far, being almost identical to the real data.

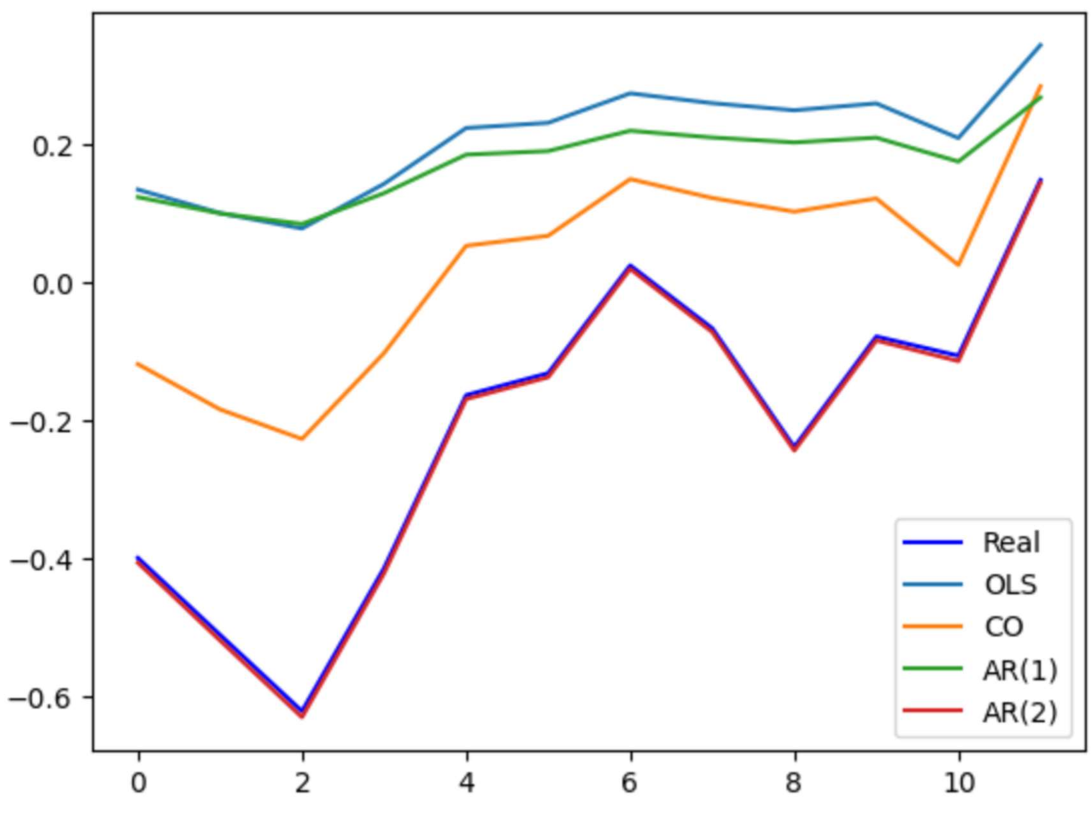

Figure 2: Real and predicted current at different time points using different model structures

Furthermore, we analysed more model structures. First not accounting for autocorrelation and then, including autocorrelation.

We analysed more complex regression models, including linear and polynomial support vector machines (SVR), a multilayer perceptron (MLP), gradient boosting (GB) and extreme gradient boosting (XGB).

The first can be seen in Figure 3. Here, the MLP and the SVR have the best results, as their predictions represent the trend accurately, which the GB and XGB methods do not. However, the error is similar to the OLS results.

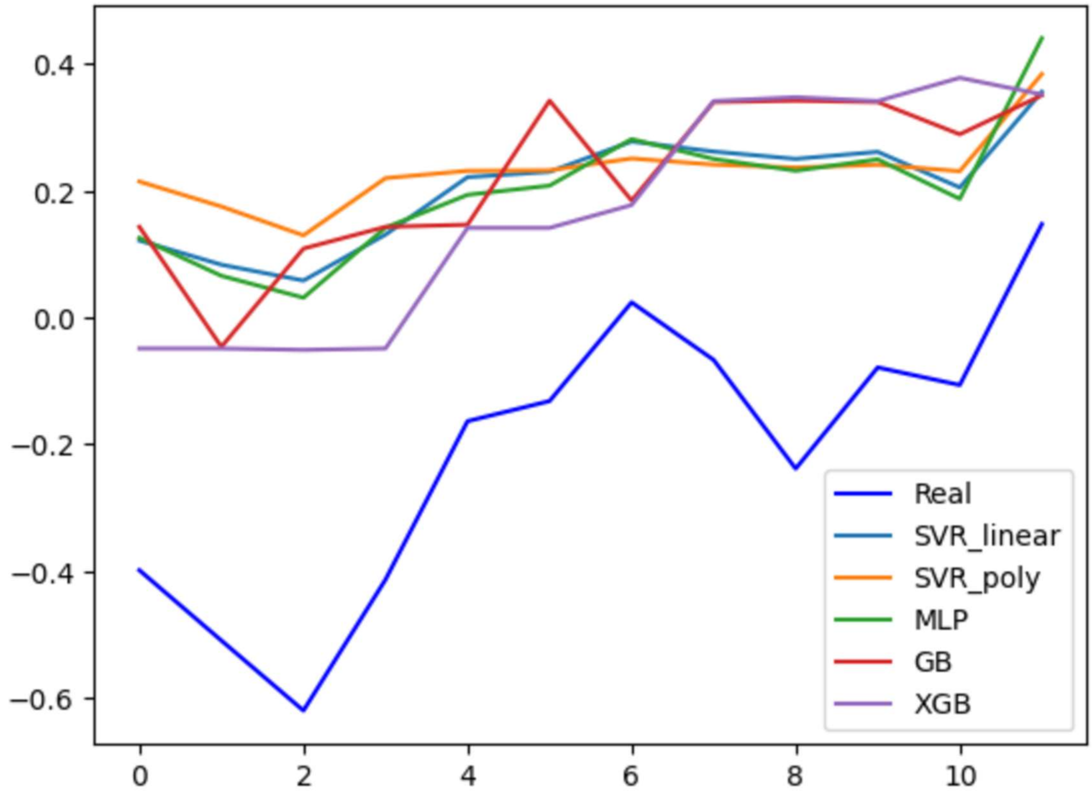

Figure 3: Real and predicted current at different time points using support vector machines, multilayer-perceptron, gradient boosting and extreme gradient boosting; not accounting for autocorrelation.

Figure 4 shows the results of the more complex models, accounting for autocorrelation. Now, the absolute error is less. However, the representation of the trends is much worse. Only the linear SVR shows results than the Cochrane-Orcutt transformation model. However, it is not closely as accurate as the linear autoregressive model, that accounted for load sums.

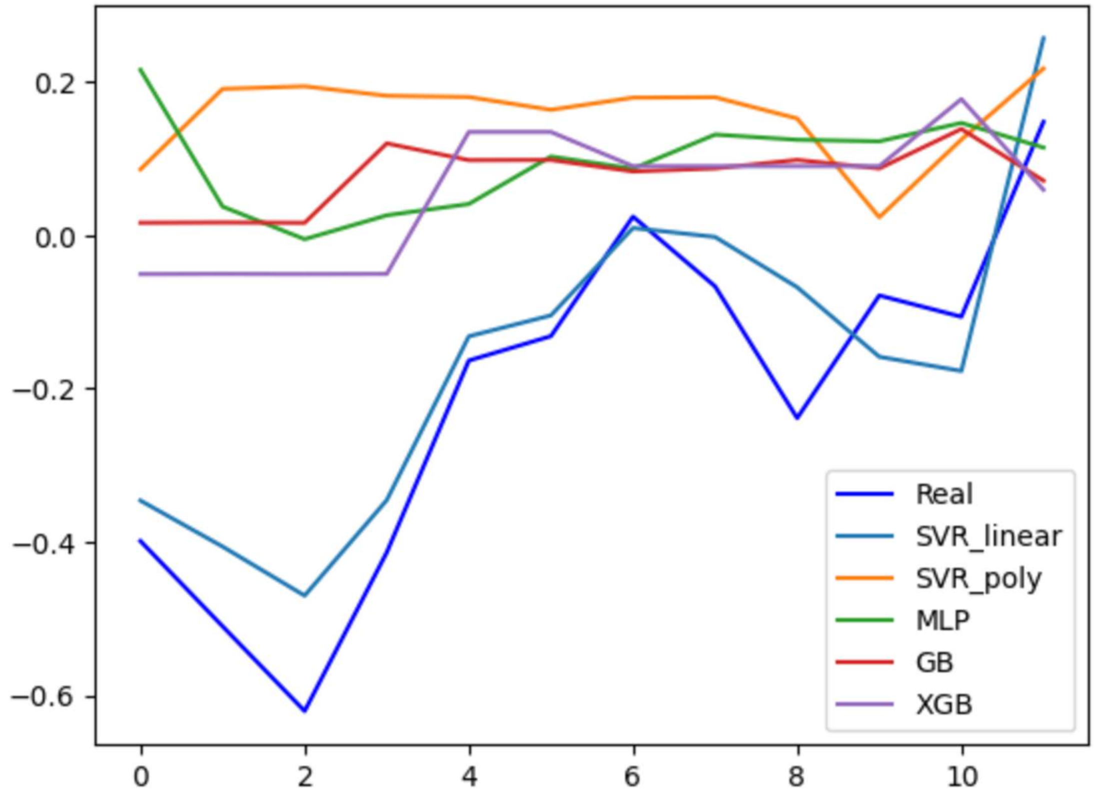

Figure 4: Real and predicted current at different time points using support vector machines, multilayer-perceptron, gradient boosting and extreme gradient boosting; accounting for autocorrelation.

Conclusion

Among the models examined, the linear autoregressive model incorporating load sums stands out as the most effective approach. The success of the autoregression model with load sums can be attributed to several factors. By accounting for autocorrelation in the data, the autoregressive model captures the temporal dependencies and dynamics inherent in the system. This allows the model to effectively learn from past observations and make accurate predictions based on the current state of the system. Including load sums as independent variables provides the model with additional information about the overall demand on the grid. This enables the model to capture the collective impact of various loads on the current flow, leading to more accurate predictions.

Moreover, the observed discrepancy in the performance of more complex models compared to linear regression, is probably a result of our very small training dataset. Complex models typically involve a larger number of parameters that need to be estimated from the data. With a small training dataset, there may not be sufficient information to accurately estimate these parameters, leading to high variance in the model's predictions. This variance can manifest as instability in the model's performance across different subsets of the data, resulting in unreliable predictions.

In contrast, linear regression, being a simpler model, has fewer parameters and lower computational complexity. It is less prone to overfitting, especially in scenarios with limited data, and can provide more stable and interpretable predictions. Additionally, the linear nature of the model makes it more resilient to noise and outliers in the data, which can be advantageous in situations where the data are noisy or subject to measurement errors.