PREDICTING DAY-AHEAD ELECTRICITY PRICES IN THE NORDIC

Søren Blatt Bendtsen, s164521, Julius Lindberg Steensberg, s193885,
Frederik Vejby Nielsen, s183709, Søren Nielsen Sardemann, s193893

42186 - Model-based Machine Learning

Abstract

This report investigates the prediction of day-ahead (DA) electricity prices in the Nordic bidding zones, particularly zone DK1, using model-based machine learning techniques. It focuses on the implementation of Linear Dynamical Systems (LDS) and Gaussian Processes (GP) to capture the complex temporal and spatial dynamics influencing electricity prices. The study utilizes data from ENTSO-E, covering a full year of hourly records for various features, including wind and solar forecasts, load forecasts, and transmission capacities. The impact of dimensionality reduction via Probabilistic Principal Component Analysis (PPCA) on model performance is also examined. Key findings indicate that the best-performing model, based on root mean squared error (RMSE), was the GP using the scikit-learn library. In terms of computational time, the LDS model showed significant benefits, especially when combined with PPCA. Additionally, a multivariate LDS was developed to predict prices in both DK1 and DK2. While this model performed reasonably well, it had difficulties predicting times when the prices in DK1 and DK2 diverged. Despite the substantial dataset size, the DA prices proved very difficult to forecast, highlighting the challenges of predicting the complex and interconnected Nordic electricity market.

1 The Nordic energy market

The price of electricity is established on the day-ahead market, where the price is set at 12:00 noon the day before for the 24 hours of the following day. This means that after 12:00, the prices for the next day are known. These day-ahead prices (DA prices) are set for each bidding zone. In Denmark, there are two bidding zones, whereas Norway, for instance, has five bidding zones. The bidding zones are connected by transmission lines, as shown in Figure 1 for the Nordic Region, which allows for trade between the bidding zones, limited by their capacity. This makes the Nordic system a highly interconnected system and the DA prices are therefore correlated. Bidding zones have equal DA prices if the transmission is not congested since the marginal producer in the most expensive bidding zone also becomes the marginal producer in the region to which power is exported, as long as there is sufficient transmission capacity.

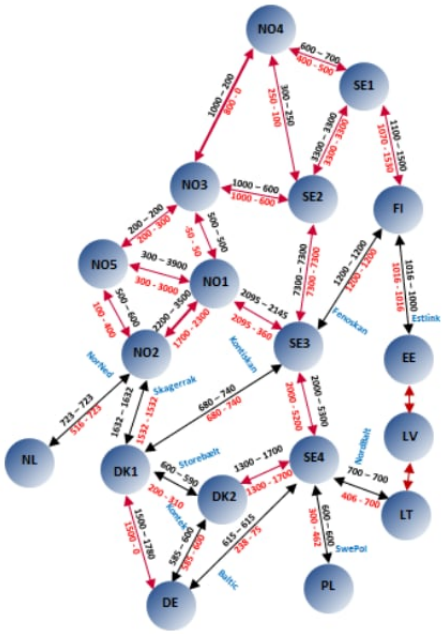


Fig. 1: Diagram of the north European bidding zones and their connections [1].

Also, the DA price for each bidding zone is influenced by the demand for electricity at the given hour for the bidding zone, the supply of renewable power generation (wind, solar, etc.), and conventional power generation, but also by the production and consumption in the neighboring bidding zones (even neighbors' neighbors) due to the interconnectivity and trade between bidding zones.

The goal of this project is to model all of these dynamics in order to provide the best prediction possible of the DA prices in the Nordic bidding zones. To simplify things a bit, the primary focus will be on forecasting the DA prices of DK1 (covering Jutland and Funen), leading to the following research question:

- **What factors influence the day-ahead prices in DK1, and how can these dynamics be accurately modeled to predict the DA-prices using model-based machine learning techniques?**

2 Data Selection

In order to model this problem, data has been obtained from ENTSO-E (European network of transmission system operators for electricity), which has a publicly available API [2]. A variety of datasets have been queried from the API and after a thorough data pre-processing part, a complete data set for the Nordic Region has been created. The following data has been gathered for each of the 12 Nordic bidding zones:

- UTC Timestamp: Date with hourly granularity. Additional features are extracted from the Timestamp such as Hour and Weekday, which are then processed with a cyclical Sine and Cosine transformation to correctly capture it's numeric importance.
- Day Ahead Price: Actual electricity price for the Day Ahead market [EUR / MWh]
- Wind Offshore: Forecast of offshore wind generation [MWh]
- Wind Onshore: Forecast of onshore wind generation [MWh]
- Solar: Forecast of solar generation [MWh]
- Total Generation: Forecast of total power generation [MWh]
- Forecasted Load: Forecast of power demand [MWh]
- Capacity to: Forecasts of power capacity in transmission lines to neighboring bidding zones [MWh]. Mostly constant, but sometimes has reductions due to maintenance which highly impacts the Day Ahead Price in connected bidding zones.

The data was scraped and prepared for the entire 2023. First of all, it was important to include data across an entire year to learn the different seasons that highly impact electricity prices. Furthermore, 2023 was a year with more "normal" prices compared to e.g. 2022 which was highly impacted by the beginning of Russia's invasion of Ukraine.

The complexity of the data, characterized by temporal changes over time and spatial connections between bidding zones, makes it very challenging to model, and predicting electricity prices is currently a very relevant topic. The expectation, however, is that we have gathered sufficient data to capture these dynamics using model-based machine learning techniques.

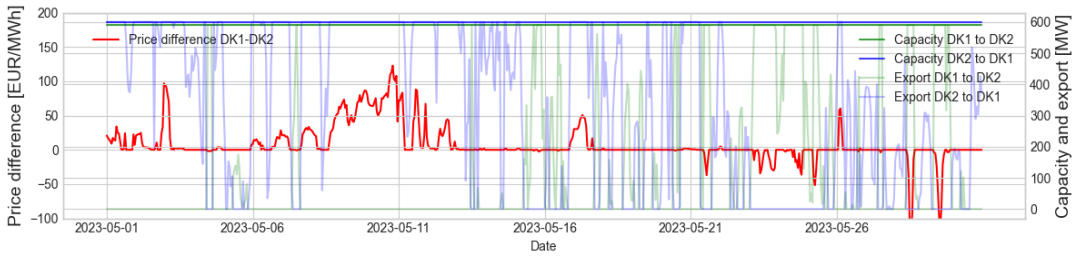
3 Data Description

The final dataset (8736,104) has 92 input features and 12 target variables with day-ahead prices for the Nordic bidding zones, although the primary focus is to predict prices in DK1 only. A multivariate LDS, however, will be implemented to predict multiple prices at once. Furthermore, each row in the dataset corresponds to a unique hour. This has been achieved by pivoting the dataset into additional features (Wind SE4, Wind NO1, etc.). This leads to a high-dimensional data set making it relevant to explore feature selection and dimensionality reduction techniques.

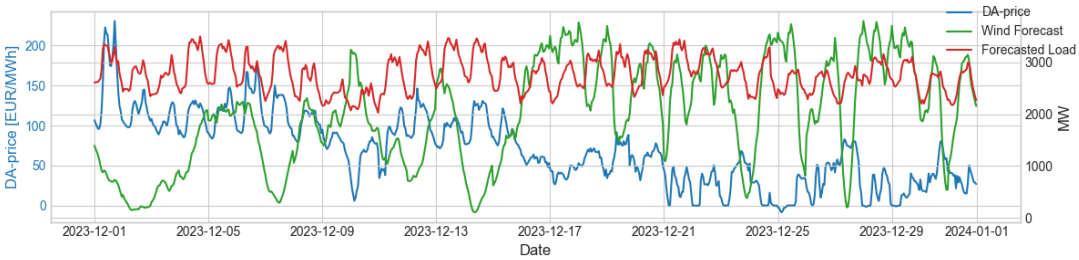
As mentioned in the Background section, two bidding zones will have equal day-ahead (DA) prices if the transmission lines between them are not congested, allowing for additional import or export. This relationship is illustrated in Figure 2a. The **red** line, plotted against the left y-axis, represents the price difference between DK1 and DK2. A value above 0 indicates a higher DA price in DK1, while a value below 0 indicates a higher DA price in DK2. On the right y-axis, the thick top **blue** and **green** lines show the forecasted transmission capacity from DK2 to DK1 and from DK1 to DK2, respectively. The light blue and light green lines indicate the actual export between the zones. Figure 2a demonstrates that when neither the light blue nor the light green line reaches maximum capacity (indicating that the transmission lines are not congested), the DA prices in DK1 and DK2 are equal. However, when the transmission line from DK2 to DK1 is congested, the DA price is higher in DK1. Conversely, when the transmission line from DK1 to DK2 is congested, the DA price is higher in DK2. The plot also indicates that forecasted capacity aligns closely with actual export, increasing their trustworthiness as features for predicting electricity prices. However, the actual exports can't be utilized in our price predictions since those are measured at least 24 hours after the prices have been announced. Initially, the goal was to predict export forecasts as well since these are not available publicly but it was decided to keep the scope of our project narrow.

In conclusion, the electricity market is heavily influenced by physical limitations in the grid, making price predictions particularly challenging. Prices are also highly dependent on the sources of production, as some types are cheaper than others. When the load is high, it often necessitates the activation of more expensive production. This dynamic is evident in Figure 2b. The plot depicts the relationship between the DA price (blue), wind forecast (green), and load forecast (red) for DK1 in December 2023. There is a clear inverse relationship between the DA price and wind forecast; when the wind forecast is high, the DA price tends to be lower, and vice versa. Additionally, daily fluctuations are evident in the load, which is typically low during the night and high during the day. DA prices generally follow this pattern because cheap renewable energy sources rarely cover the total load at its peak, necessitating the activation of more expensive conventional generators.

In conclusion, forecasts of capacity, wind, and load are likely the most critical features for predicting electricity prices in 2023. However, due to the complexity and interconnected nature of the system, we have also included solar and total generation in our analysis. While these features provide valuable insights, incorporating weather data could further improve the



(a) Day-ahead price dynamics between DK1 and DK2 for May 2023



(b) Relationship between the day-ahead (DA) electricity price, wind forecast, and load forecast for DK1 in December 2023

predictions, but such data was not acquired for this project.

4 Modelling Approach

The models that will be used to predict the DA prices are based on linear dynamical systems (LDS) and a Gaussian Process (GP).

4.1 Linear Dynamical System

A general class of models for time-series data is the class of State-Space Models (SSM's), which stems from the same idea as the auto-regressive models, where it is assumed that each observation is a linear combination of M previous observations. However, state-space models differ by instead introducing an extra layer of hidden variables from which the observations are generated, as shown in figure 3.

A state-space model can be described by two probabilities: the transmission probability and the emission probability. If we assume linear Gaussian distributions for the two probabilities, we end up with the sub-class of Linear Dynamical Systems (LDS).

The simplest LDS only seeks to learn the latent states and relations based on the initial values. However, we also assume a dependency on i.e. the production data. To account for this, we include an input dimension to the latent states at each time t. This results in a transmission probability given by

$$\mathbf{h}_t \sim \mathcal{N}(\mathbf{h}_t|\mathbf{B}_1\mathbf{h}_{t-1} + \mathbf{B}_2\mathbf{h}_{t-2} + \mathbf{W}\mathbf{x}_t, \mathbf{R}). \quad (1)$$

The corresponding emission probability for the model is given by

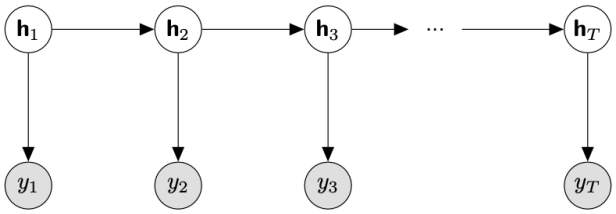


Fig. 3: Generic state-space model [3]

$$y_t \sim \mathcal{N}(y_t|\mathbf{h}_t, \sigma^2). \quad (2)$$

Using the described probabilities, we end up with the generative process described in algorithm 1 with the corresponding probabilistic graphical model in figure 4.

Algorithm 1 Univariate linear dynamical system

Draw transmission coefficients $\beta \sim \mathcal{N}(\boldsymbol{\beta}|\mathbf{0}, \lambda\mathbf{I})$
Draw transmission noise variance $\tau \sim \text{HalfCauchy}(1)$
Draw emission noise variance $\sigma \sim \text{HalfCauchy}(3)$
Draw input coefficient $\boldsymbol{W} \sim \mathcal{N}(\boldsymbol{W}|\mathbf{0}, \lambda\boldsymbol{I})$
Draw first latent variable $h_1 \sim \mathcal{N}(h_1|0, 1)$
Draw second latent variable $h_2 \sim \mathcal{N}(h_2|\beta_1 h_1, 1$
for each time $t \in \{3, \ldots, T\}$ **do**
    Draw latent variable
    $h_t \sim \mathcal{N}(h_t|\beta_1 h_{t-1} + \beta_2 h_{t-2} 0\boldsymbol{W}\boldsymbol{x}_t, \tau)$
    Draw observation $y_t \sim \mathcal{N}(y_t|h_t, \sigma^2)$
end for

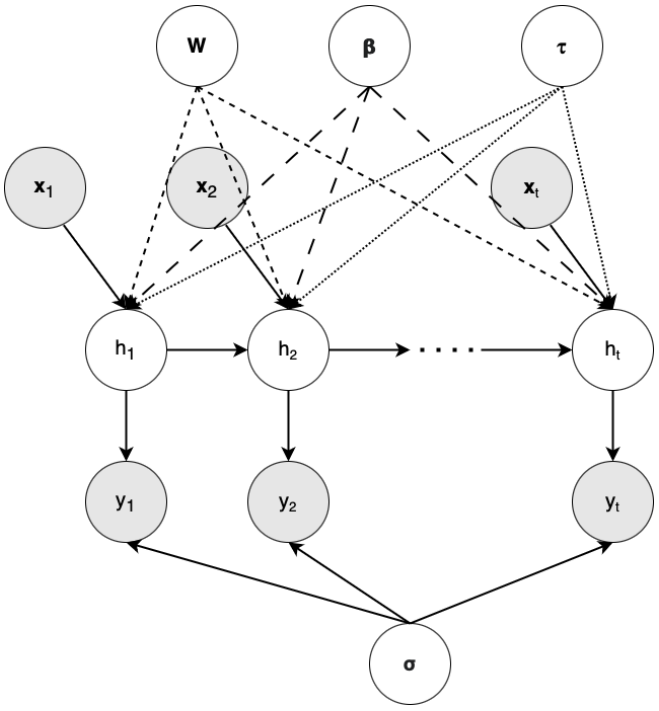


Fig. 4: Probabilistic graphical model describing the univariate LDS for one bidding zone.

As stated in section 1, there is an interaction between neighbouring bidding zones, which influence the day-ahead price in the zones. Thus, the model is expanded into a multivariate linear dynamical system to capture the correlation between neighbouring bidding zones. The primary difference from the previous model for a single bidding zone is that we group the day-ahead prices for all investigated bidding zones into one observation vector, $\mathbf{y}_t = (y_t^1, \ldots, y_t^M$, with M being the number of investigated bidding zones.

For a multivariate linear dynamical system, the transmission and emission probabilities become:

$$\begin{aligned} \mathbf{h}_t &\sim \mathcal{N}(\mathbf{h}_t|\mathbf{B}_1\mathbf{h}_{t-1} + \mathbf{B}_2\mathbf{h}_{t-2} + \mathbf{W}\mathbf{x}_t, \boldsymbol{\Sigma}) \\ \mathbf{y}_t &\sim \mathcal{N}(\mathbf{y}_t|\mathbf{h}_t, \sigma^2\mathbf{I}). \end{aligned} \tag{3}$$

The essential difference is that we now have a covariance matrix for the transmission probabilities. In order to accurately incorporate some degree of correlation between the bidding zones, we need the covariance matrix $\boldsymbol{\Sigma}$ to be full.

As with most probabilistic models, the main challenge is defining priors for the model parameters. As explained in the lecture 6 notebook [4], the inverse-Wishart distribution, which is the conjugate prior for a multivariate normal distribution, is not well-suited for modern Bayesian computational methods.

Instead, we follow the standard approach of using an LKJ (Lewandowski-Kurowicka-Joe) distribution as the prior for modelling the covariance matrix, given by

$$\boldsymbol{\Sigma}_{\text{lower triangular}} = \text{diag}(\boldsymbol{\tau})\boldsymbol{\Omega} \tag{4}$$

$\text{diag}(\boldsymbol{\tau})$ denotes a diagonal matrix with the diagonals given by the vector $\boldsymbol{\tau}$. Thus, $\boldsymbol{\Omega}$ can be regarded as the correlation matrix and $\boldsymbol{\tau}$ as a vector of coefficient scales.

The multivariate linear dynamical system is almost similar to the probabilistic graphical model in figure 4, with the exception that $\boldsymbol{h}$ and $\boldsymbol{y}$ are now vectors and the transmission noise is now given by a full matrix, $\boldsymbol{\Sigma}$.

4.2 Gaussian Processes

Gaussian Processes (GPs) is a powerful non-parametric, probabilistic technique that is especially useful for time series forecast problems involving continuous data with complex relationships [5]. That makes GP a relevant technique for this project. In this project, a simple GP model was created comparing different Kernel (Covariance) Functions.

Although GPs are non-parametric models, there are still important choices to be made and hyperparameters to be tuned in order to create a good model. One of these is the choice of kernel, which defines the model behaviour by measuring pairwise similarities between data points. For this project, a combination of the Squared Exponential (SE) and Periodic kernels were chosen. Their mathematical formulas are shown in Equation 5 and 6 below, and the combination basically works well for this specific project as the SE kernel captures the smoothness of the data making the model generalize well, while the Periodic kernel captures well the periodicity of the data which is quite strong when it comes to electricity prices in markets with high renewable penetration.

$$k_{SE}(x, x') = exp\Big(-\sum_{d=1}^{D}\frac{(x_d - x'_d)^2}{2l^2}\Big) \tag{5}$$

$$k_{PER}(x, x') = h^2 exp\Big(-\frac{1}{2l^2}sin^2\Big(\frac{\pi}{p}\sum_{d=1}^{D}(x_d - x'_d)\Big)\Big) \tag{6}$$

The SE kernel has the positive length-scale parameter l which determines the wiggles of the function [6]. A very low $l = 0.0001$ was chosen as it gave the most smooth function where nearby observations covary more. The Periodic kernel has the additional parameters h and p as well as l, which defines the amplitude and periods respectively. The parameters for the Periodic kernel were chosen to $h = 1, l = 0.05, p = 2$. With the use of marginal likelihood optimization, it is possible to find the optimal parameter choice. However, this was not done for this project, as the original chosen parameters turned out to perform very well when making predictions. When

making predictions for y_* given a new input $\mathbf{x}_*$, the distribution over y_* is given by:

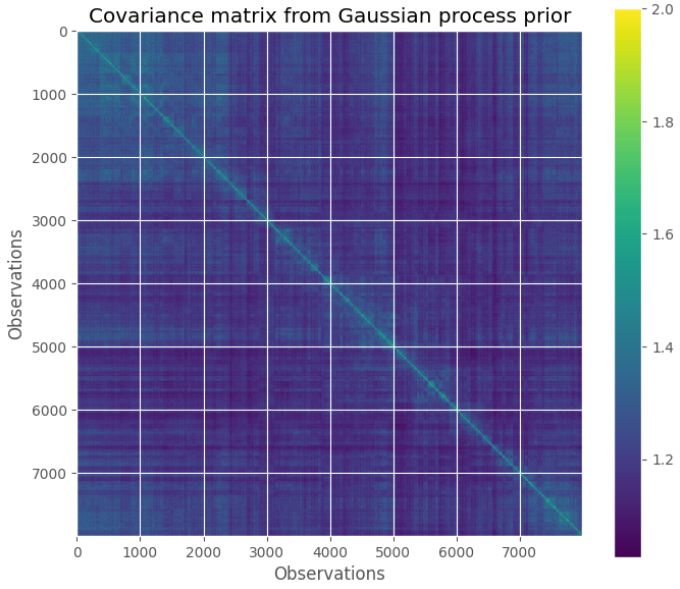
$$p(y_*|\mathbf{y}, \mathbf{x}_*, \mathbf{X}) = \mathcal{N}\Big(y_*|\mathbf{k}_*^T(\sigma^2\mathbf{I}+\mathbf{K})^{-1}\mathbf{y}, k_{**}+\sigma^2-\mathbf{k}_*^T(\sigma^2\mathbf{I}+\mathbf{K})^{-1}\mathbf{k}_*\Big) \tag{7}$$

$\mathbf{K}$ is the covariance function based on training data, $\mathbf{k}_*$ is based on the combination of training and test data and k_{**} is based on test data. $\mathbf{K}$ and k_{**} are visualized in Figure 5. The periodic structures are visible as repeating blocks or grid-like patterns, while the smooth variations are indicated by the gradual change in color intensity away from the diagonal. This confirms that the GP is capturing both smooth trends and periodic behaviors in the data, as intended by the choice of kernels.

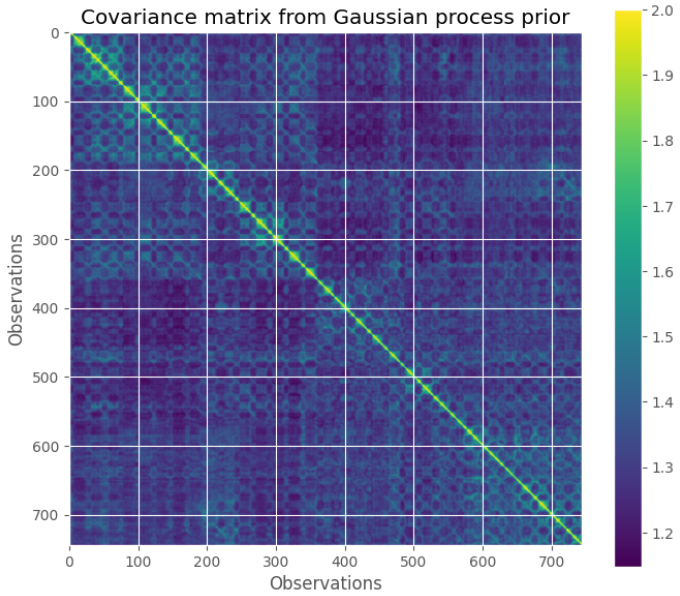
As a final remark, two GPs with common kernels and hyperparameters will be implemented. A GP based on the lecture notebooks and a GP based on a scikit-learn implementation. The performance of the GPs and the temporal models are compared in the Results section.

5 Dimensionality Reduction with PPCA (ARD)

The dataset used in this project is high-dimensional. Therefore, a probabilistic principal component analysis (PPCA) with Automatic Relevance Determination (ARD) is utilized to reduce the dimensions of the input data to investigate how differently the models perform on the original dataset and the PPCA projections, respectively. This assessment is based on runtime vs model performance. Figure 6 shows the PGM of the PPCA with ARD, while Algorithm 2 shows the generative process. Here, $\mathbf{W}$ represents the linear transformation that maps the latent projections $\mathbf{z_n}$ to the observed data $\mathbf{x_n}$ (similar to principal components). σ is the noise variance across all observations, while α are the precision parameters, which tells us how relevant the dimensions of our $\mathbf{W}$ matrix is. Thus, ARD is applied to the PPCA to ensure that the most relevant principal components are emphasized. The PPCA is run with 10 latent dimensions, i.e., the input data is projected into 10 dimensions. The posterior α mean values obtained are [2.9, 3.2, 3.0, 2.5, 3.0, 1.3, 3.0, 2.4, 3.2, 2.4]. High values imply a small variance, meaning the model believes the weight $\mathbf{w}_k$ should be close to zero, and vice versa. In this case, the α's are quite similar, indicating a similar relevance of the components, although $\mathbf{w}_6$ appears to be the most relevant component. Furthermore, $\mathbf{z}$ and $\mathbf{W}$ can be used to reconstruct our input data in the original input space by reversing the projection as $\mathbf{z}.\mathbf{W}^T$. The reconstruction comes with an error and this error can be used to assess the performance of the projections. This can be visualized by



(a) Covariance matrix $\mathbf{K}$ - train data



(b) Covariance matrix k_{**} - test data

Fig. 5: Kernel combination of Squared Exponential and Periodic

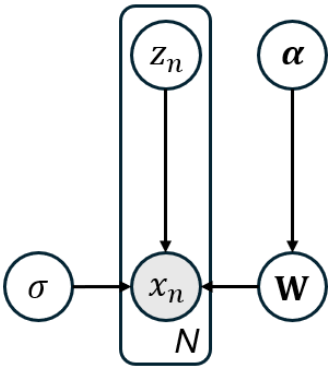


Fig. 6: PGM - PPCA (ARD)

Algorithm 2 Generative Process for PPCA with ARD

Draw noise variance $\sigma \sim \text{HalfCauchy}(1)$
for each latent dimension $k \in \{1, \ldots, K\}$ **do**
    Draw $\alpha_k \sim \text{Gamma}(1.0, 1.0)$
end for
Draw weights $\mathbf{W} \sim \mathcal{N}\left(0, \alpha^{-1}\mathbf{I}\right)$
for each data point $n \in \{1, \ldots, N\}$ **do**
    Sample latent projection $z_n \sim \mathcal{N}(0, \mathbf{I})$
    Sample observation $x_n \sim \mathcal{N}(\mathbf{W}z_n, \sigma\mathbf{I})$
end for

plotting a reconstructed input feature vs the original, which has been done for the Forecasted Load as shown in Figure 7. The reconstructed Forecasted Load appears to be satisfactory, raising expectations that the projections will also be reliable. Therefore, in the following section, both the original data and the projected data will be used as inputs for the univariate temporal model and two GP models. This approach aims to determine whether a beneficial balance between runtime efficiency and model performance can be achieved using dimensionality-reduced data.

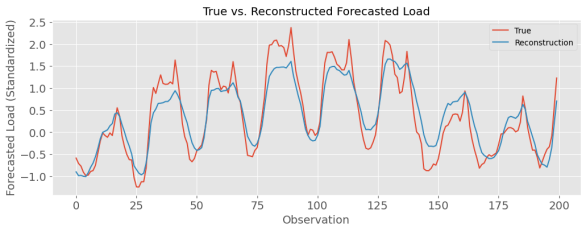


Fig. 7: PPCA reconstruction vs the original normalized forecasted load (first 200 obs.)

6 Results

The models described in Sections 4 are trained on the first 11 months of data, and their performances are evaluated using a test set covering December 2023. While the models showed reasonable performance when trained on just 1-2 months of historical data, their predictions improved significantly with 11 months of training data at the expense of a larger computational load.

6.1 GPs & Univariate LDS

First, the performance of the two GP models and the univariate temporal model will be examined. Table 1 and Figure 8 present a comparative analysis of the model performances in predicting the day-ahead electricity prices in DK1.

Table 1 provides an overview of the Root Mean Square Error (RMSE) for the normalized DA-price and runtime for each model. The three models evaluated are the Temporal Model, which has an RMSE of 0.797 and a runtime of 18 minutes; the GP from scikit-learn, which shows the best performance with an RMSE of 0.504, although it has a longer runtime of 106 minutes; and the Self-made GP, which has the highest RMSE of 0.999, with a runtime of 13 minutes. Table 1 demonstrates a clear correlation between runtime and the accuracy of predictions. While the GP from scikit-learn provides the most accurate predictions, it requires significantly more computational time compared to the other models. The Temporal Model, on the other hand, strikes a balance between accuracy and runtime efficiency. Figure 8 illustrates the predicted

Model	RMSE	Runtime
Temporal	0.835	18 minutes
GP from scikit-learn	0.504	106 minutes
Self-made GP	0.999	13 minutes

Table 1: Model RMSE on the normalized DA-price and runtime

versus actual prices for the three models. The plot reveals that the GP from scikit-learn closely follows the actual prices, confirming its superior accuracy as indicated by the RMSE values. The Temporal Model also shows a good fit but with slightly higher deviations from the actual prices. The Self-made GP model demonstrates larger discrepancies, reflecting its higher RMSE.

In summary, the GP from scikit-learn is the most accurate model, although it still has some significant deviations from the actual prices. A noteworthy benefit of the GP from scikit-learn is that it generally follows the price trend very well. While the magnitude might be slightly off, it often accurately captures when the day-ahead price will increase or decrease throughout the day. This is highly valuable in real-life use cases because consumers are often more concerned with the relative differences in price throughout the day for activities such as vehicle charging rather than the actual price level. On the other hand, the Temporal Model has more fluctuations, making it harder to discern the price trend over the day. These

fluctuations are partly due to the positive and negative values for β_1 and β_2, where the inclusion of more β-values might had given a less fluctuating prediction.

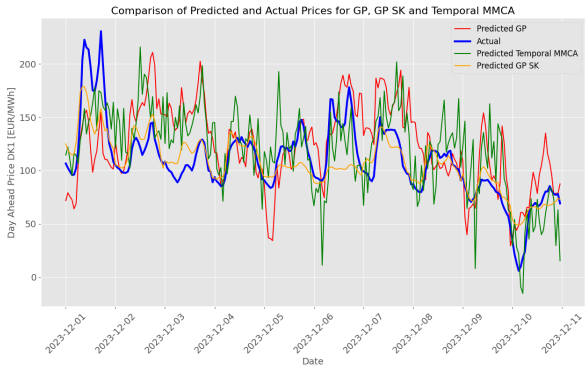


Fig. 8: Predicted vs actual prices (all models)

6.2 GPs & Univariate LDS (PPCA data)

Figure 9 presents a comparison of the predicted versus actual prices using the original data and the projections from Probabilistic Principal Component Analysis (PPCA). This analysis focuses on evaluating the impact of dimensionality reduction on the models' performances.

Since the covariance matrices in the GP are dependent on the number of observations, reducing the number of features did not at all reduce the computational time. So, using PPCA data for the GPs only worsened the predictions.

Conversely, the temporal model benefited from the feature reduction, as evidenced by the runtime in Table 2. There was a runtime reduction of 22%, although this was accompanied by a slightly higher RMSE. Overall, while the slight increase in RMSE indicates a small drop in prediction accuracy, it could definitely be argued as a decent trade-off given the benefits of faster computation and reduced model complexity when combining the temporal models with dimensionality-reduced data.

Model	RMSE	Runtime
Temporal	0.835	18 minutes
Temporal PPCA	0.867	14 minutes
GP from scikit-learn	0.504	106 minutes
Self-made GP	0.999	13 minutes

Table 2: Model performance and runtime with the temporal PPCA.

6.3 Feature importance

It is valuable to identify the most useful features for predicting DK1 prices. Such analysis will be conducted on the Gaussian

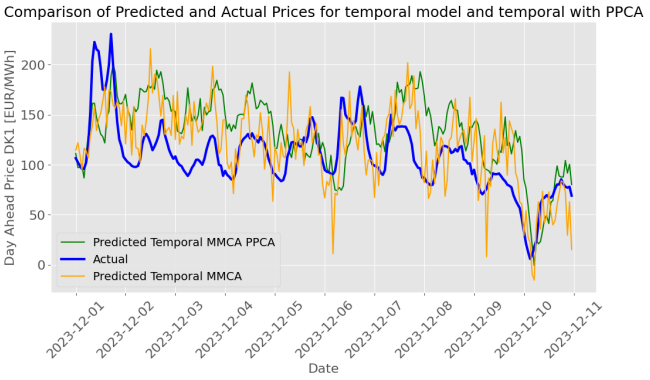


Fig. 9: Predicted vs actual prices using original data and PPCA projections for the temporal model.

Process model from scikit-learn, as it demonstrated the best performance. Figure 10 showcases a bar chart that ranks the importance of the features in predicting day-ahead electricity prices. Each bar represents a feature, with its height indicating the degree to which predictions vary when that feature is perturbed by 0.01 of its standard deviation while holding other features constant. The importance measures are comparable because all input features are standardized.

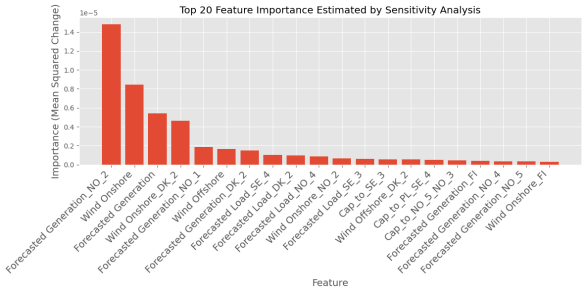


Fig. 10: Top 20 most important features for the GP SK model found by perturbation of the standardized X test data

The features without bidding zones in their name refer to the specific bidding zone being modeled. For example, the second most important feature, 'Wind Onshore,' represents the onshore power generation forecast for DK1. It is reasonable that onshore, offshore, and total forecasted generation in DK1 heavily affects the price prediction. Interestingly, the neighboring bidding zones NO2 and DK2 (see Figure 1), along with their respective forecasted generation and onshore forecasts, are also highly important. For the forecasted generation in NO2, it makes sense since DK1 is heavily affected by the potential hydropower generation in the Norwegian bidding zones. In times of low wind generation in Denmark, DK1 imports cheap and stable hydropower from Norway.

Additionally, some of the capacity features like 'Cap to SE3' affect the price. This means that the price is influenced by

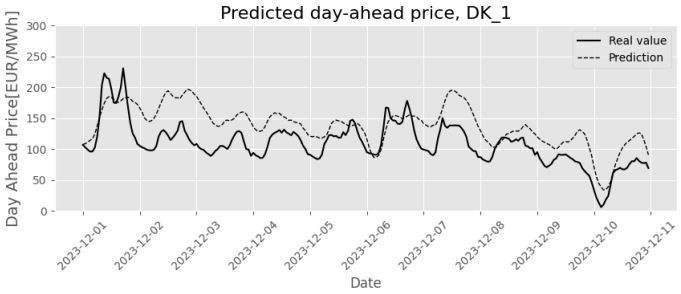
how much capacity is available for transmission between SE3 and DK1. This is reasonable because DK1 has the possibility to import stable nuclear power from SE3 or cheap hydro power which has been transported down from SE2 and SE1 and exported from SE3 to DK1. However, in times of low capacity on the transmission line, this possibility diminishes which affects the DK1 DA prices.

Lastly, in a more general context, although the importance is much lower, Figure 10 visualizes that even features as far away as in the FI or NO4 bidding zones affect the price prediction in DK1, as seen with the features 'Forecasted Load NO4' and 'Forecasted Generation FI.' The fact that these two features have an effect on the price prediction in DK1 shows the complexity and interconnections between the bidding zones, making the day-ahead price prediction a difficult challenge.

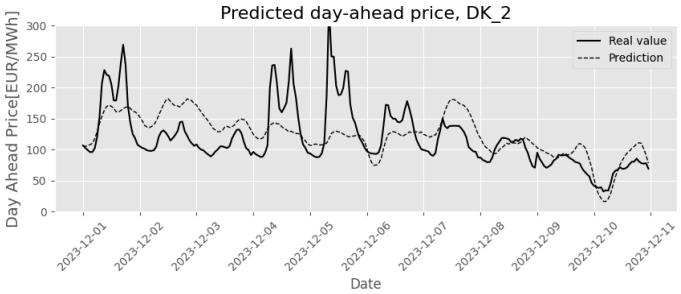
For comparison, one can inspect Figure 12 in the Appendix which highlights some of the distributions of latent weights in the temporal model. It is observed that among the weights plotted, the most influential features (based on the scaling on the x-axis, since all the features are normalized) are the forecasted generation in DK1, forecasted generation in NO2, and forecast of onshore wind in DK1. This aligns completely with Figure 10. On the other hand, the capacity on the transmission line to the Netherlands is barely given any weight in Figure 12.

6.4 Multivariate LDS (DK1 & DK2)

Due to the interconnectivity of the power system, electricity prices are best predicted by viewing the system of bidding zones holistically rather than treating them as independent targets. Although multioutput predictions are slightly beyond the scope of this project, it was decided to anyway include an investigation of the performance of a multivariate LDS, as described at the end of Section 4.1, in predicting two highly correlated electricity prices: DK1 and DK2. Figure 11 shows the price predictions for DK1 and DK2, while Table 3 presents the RMSE and runtime. The multivariate LDS still lags behind the Gaussian Process (GP) from scikit-learn in terms of RMSE. However, the predictions for DK1 have improved compared to previous temporal models, despite the combined runtime of 71 minutes for predicting two zones being substantially higher than the 18 minutes required by the univariate temporal model. Figure 11 illustrates that the actual prices of DK1 and DK2 are very similar for most hours. Likewise, the predictions from the multivariate temporal model follow similar patterns, although the predicted prices for DK1 are higher than those for DK2. This matches real-life expectations, where average prices in DK1 are typically a bit higher than in DK2, making December 2023 an unusual case. This could explain why the model tends to overshoot the DK1 predictions for December 2023. Nonetheless, similar prediction patterns indicate that the multivariate temporal model places significant emphasis on the correlation between DK1 and DK2 prices. Therefore, it seems reasonable that the model struggles to forecast price peaks, where the prices of DK1 and DK2 suddenly diverge (see December 4th and 5th in Figure 11b).



(a) Multivariate temporal prediction for DK1



(b) Multivariate temporal prediction for DK2

Fig. 11: Multivariate temporal model prediction for DK1 and DK2.

Model	RMSE	Runtime
Temporal	0.865	18 minutes
Temporal PPCA	0.867	14 minutes
GP from scikit-learn	0.504	106 minutes
Self-made GP	0.999	13 minutes
Multivariate temporal DK1	0.858	71 minutes
Multivariate temporal DK2	0.786	

Table 3: Model performance and runtime now with multivariate temporal model.

7 Conclusion

In this report, the prediction of day-ahead (DA) electricity prices in the Nordic bidding zones was explored, focusing on zone DK1. Various model-based machine learning techniques were implemented and compared, including Linear Dynamical Systems and Gaussian Processes.

The electricity market's complexity, driven by physical grid

limitations and diverse production sources, makes accurate price predictions challenging, although a variety of features for our models, such as wind and load forecasts, transmission capacities, and solar generation were utilized. High loads necessitate the activation of more expensive production methods, impacting the DA prices. On the contrary, an inverse relationship between DA prices and wind forecasts is evident, and the DA prices follow the daily load fluctuations.

The periodic nature of electricity price development made the GP models with a sum of SE and periodic kernels highly suitable for making predictions. The GP from the scikit-learn library demonstrated the best performance in terms of RMSE, but it required significant computational resources.

The LDS models, particularly when combined with Probabilistic Principal Component Analysis (PPCA), showed considerable advantages in computational efficiency. However, dimensionality reduction did not improve the runtime of the GPs since most of the computational effort lies in computing the covariance matrix K, whose dimensions are unaffected by the reduction of features.

The multivariate LDS model, predicting prices for both DK1 and DK2, achieved reasonable accuracy. However, the high correlation between the real prices of DK1 and DK2 made it difficult to train a multivariate LDS model that could also predict time intervals where the DA prices of DK1 and DK2 diverged. An extension of our input data with more relevant information could likely improve the results.

Overall, our study underscores the importance of selecting appropriate features and the trade-offs between model accuracy and computational efficiency. Future work could explore the incorporation of additional data sources, such as weather data and hydro reserve capacities, and further refine the models to improve their predictive capabilities.

By addressing these challenges, we can better understand the dynamics of the Nordic electricity market and develop more accurate and efficient models for price prediction. This understanding is crucial for optimizing energy trading and enhancing the reliability of the electricity grid.

8 Work contribution

All group members have contributited equally to all aspects of the project.

	Søren B	Søren S	Julius	Frederik
Problem definition	25%	25%	25%	25 %
Model development	25%	25%	25%	25 %
Data analysis	25%	25%	25%	25 %
Report writing	25%	25%	25%	25 %

9 References

[1] Montel Group, "Diagram of the north european bidding zones and their connections," .

[2] Entsoe, "Entsoe-py," .

[3] Rico Krueger and Filipe Rodriquez, "Lecture 06 - temporal models," *42186 Model-based Machine Learning*, 2024.

[4] Rico Krueger and Filipe Rodriquez, "Week 7 - temporal models - part 4," JuPyTer Notebook.

[5] Geeks For Geeks, "Gaussian processes," .

[6] David Duvenaud, "The kernel cookbook: Advice on covariance functions," .

Appendix

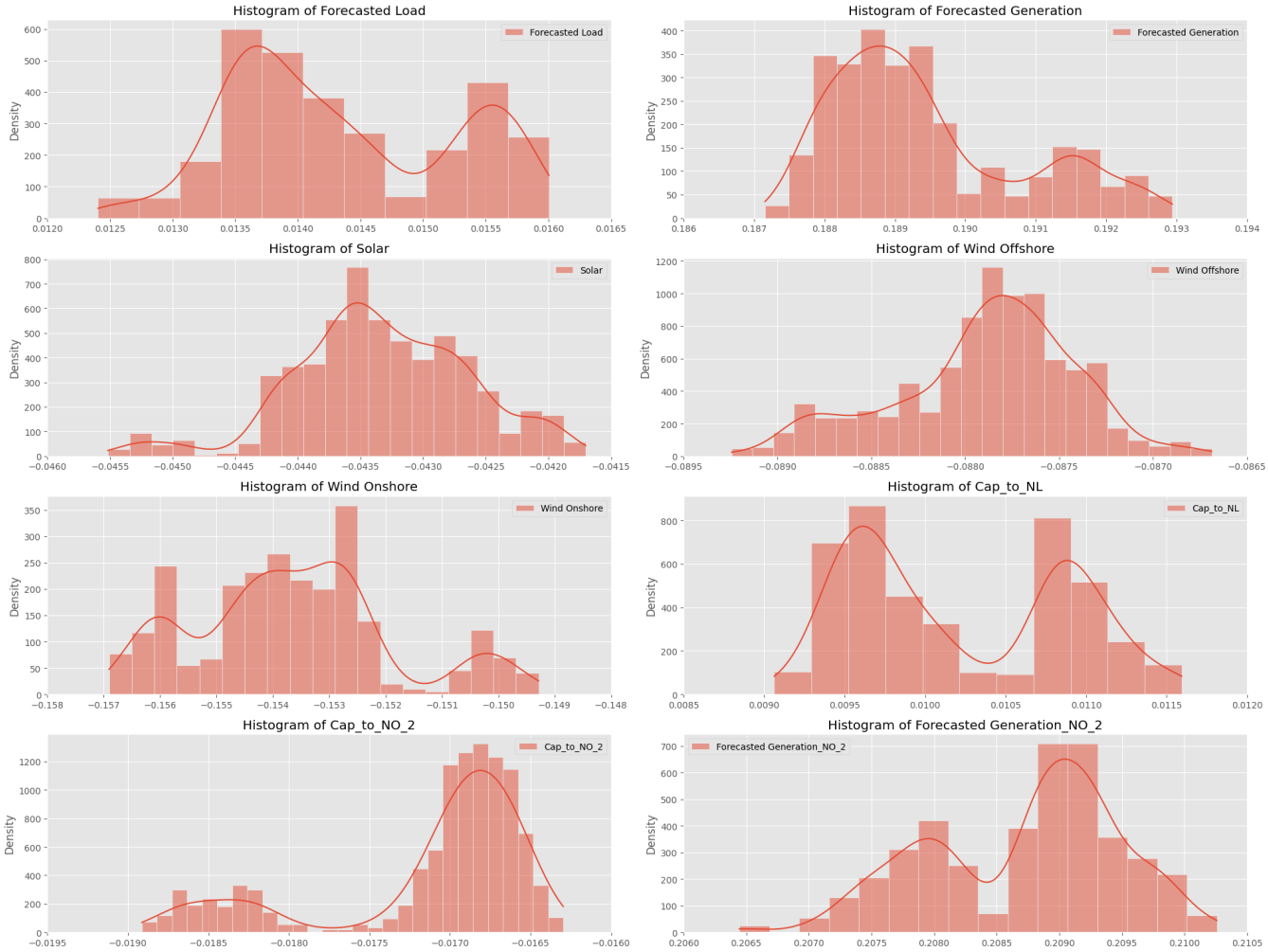


Fig. 12: Distribution for some of the weights in the temporal model