Feature Selection for Electricity Market Data through Feature Extraction

Domenic Eric Planert, Tom Bender, Florian Rippstein, Stefan Klaiber
Cognitive Energy Systems
Advanced System Technology Branch AST of Fraunhofer IOSB
Ilmenau, Germany
Correspondence: domenic.planert@iosb-ast.fraunhofer.de

***Abstract*—This paper investigates how well different combinations of data pre-processing and feature selections perform for filtering and selecting time series data for day-ahead electricity price forecasting with neural networks. We compared these against the selection of a brute force trying all combinations. Publicly available data on the electricity generation capacity of individual electricity producer groups as well as the grid load are utilized as exogenous time series to forecast the electricity price. The pre-processing methods used include signal processing methods and processing according to the Box-Jenkins model. In the feature selection methods, all three typical areas - consisting of filter, wrapper and embedded methods - are covered and examined. Additional noise time series are included in the feature selection to evaluate their robustness. The final analysis shows the weaknesses of individual methods and that a general solution is not possible due to the diverse structure of the different approaches.**

***Index Terms*--Feature Selection, Market Price Analysis, Electricity market**

I. INTRODUCTION

Electricity consumers, especially industrial companies, aim to minimize their electricity procurement costs and take advantage of procurement flexibility to benefit from fluctuating electricity prices. To plan their consumption and take advantage of this flexibility, they need a forecast of electricity prices, which is usually calculated using data-driven models. In particular, renewable energy sources, which have been greatly expanded in recent years, have a major impact on the price of electricity, as they have low electricity production costs, but their generation is highly dependent on external factors, especially weather conditions. The price of electricity depends on multiple other factors, such as generation and consumption. For a good electricity price forecast, it would make sense to include all relevant factors in the forecasting process. We are investigating which data is useful and how to select it. Reducing the data to the most crucial information is imperative, as it enhances the accuracy of the forecast, prevents overfitting of neural networks, and improves the interpretability of the models [1]. It is important to note that not all input data has a positive effect on the forecasting process. By removing irrelevant or redundant data, the signal-to-noise ratio is improved, which can lead to more accurate and robust forecasts [2]. The benefits of this approach are not limited to the forecasting process; the efficiency of neural networks is also enhanced, as the complexity of the model is reduced, and the training time is decreased. The selection of features is a crucial aspect, not only for the performance of the final model, but also for the subsequent application and maintenance of the neural network [1]. In this study, a range of feature selection (FS) methods from the three areas of wrapper, filter and embedded methods are considered. These are combined with various feature extraction (FE) techniques, the purpose of which is to extract characteristic features from the individual variables that contain concentrated information from the time series and thus provide a deeper insight. The following feature extraction techniques are examined: Signal transformations and processing according to the Box-Jenkins-method [3]. The objective is to evaluate the calculated feature sets against a brute force approach that calculates all feature combinations on three different neural networks to find the optimum. The various methods are analyzed in this paper, and in the main part of the study, the individual combinations are evaluated and compared with the brute force method. The described structure of the analysis is summarized in Fig. 1.

II. RELATED WORK

The forecasting of electricity prices is an issue that has been the focus of considerable attention and development over time [4], with initial predictions being made based purely on price and demand values [5]. The ARIMA process and wavelet approaches have been recognized as established methods in this field [6]. Recently, deep neural networks have become a subject of interest in this area [7]. Other work has emphasized that pre-processing is just as important as the correct model selection [8]. In current research, the addition of exogenous variables has also been explored. For example, a significant improvement was observed in forecasting with an extended N-BEATS model [9]. Exogenous variables offer an advantage in the transfer learning of neural networks in the use case of electricity price forecasting [10]. In scenarios where exogenous variables are incorporated, the process of feature selection becomes imperative for the purpose of identifying the most relevant variables and eliminating those that are interfer-

979-8-3503-8174-0/24/$31.00

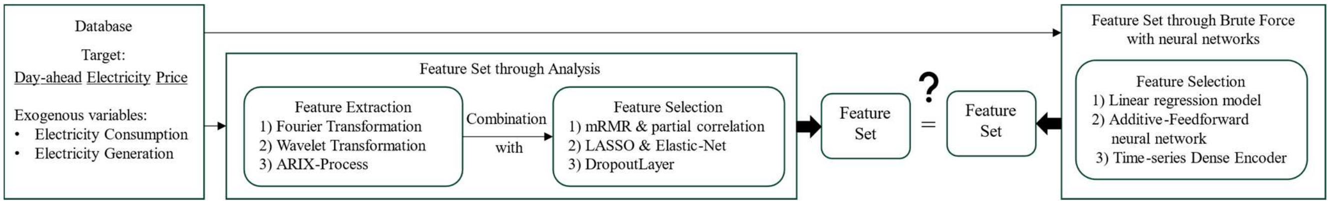

Figure 1. Database and investigated methods of feature extraction and selection for day-ahead electricity price forecasting

ing. In the study [11], a contemporary modelling scheme, known as the LSTM, is used in conjunction with methodologies for feature selection, to select from the available dataset. In addition to the extant feature selection approaches, there are two studies on a more recent embedded method. The neural network is changed such that it determines itself which features are used. For this purpose, an additional layer, which is called Dropout-Layer in the paper, is added at the beginning of the neural network, which can switch individual inputs on or off. In the work [12] this is achieved by using a Bernoulli distribution with a trainable vector that contains the individual probabilities, which is learnt through back-propagation and the error function is adapted to this learning process. This type of feature selection can recognize not only linear dependencies, but also non-linear relationships. Comparable approaches have been developed with so-called stochastic gates [13] and for Bayesian deep neural networks [14]. The three approaches share the same goal and utilize Bernoulli distributions but employ different strategies. The application of the wavelet transformation (WT) for direct forecasting was examined in [15] and [16]. The WT decomposes the time series into its components, and the forecast is then constructed using these subcomponents. The networks and methods used in the two studies differ. In both cases, the predicted individual wavelet information is summarized by the inverse transformation to form the complete time series, and an information gain can then be determined through the transformation.

III. DATA SOURCE

The data utilized in this study includes different electricity generation data, grid load and the day-ahead price from Germany in 2019. All time series have a resolution of 15 minutes, except for the price, which is set by the market on an hourly basis. The price time series is upsampled to also have a 15-minute resolution. The data is downloaded from [16]. TABLE I. lists and numbers the exogenous variables used, in addition four disturbance variables are included. The randomized disturbance variables are used to evaluate the robustness of the feature selection, with numbers 16 and 17 standing for white noise and numbers 18 and 19 being a random walk [17]. Exogenous data 3-5 have additional forecast values.

IV. METHODOLOGY

Definition: A time series Y is a sequence of equidistant values over time (1). X is a matrix of multiple time series (2).

$$Y = (y_1, \dots, y_t, \dots, y_T) \text{ with } t \in (1, \dots, T) \tag{1}$$

$$X = (Y_1, \dots, Y_p, \dots, Y_P) \text{ with } p \in (1, \dots, P) \tag{2}$$

Y is used as symbol for the target time series and X for the collection of the exogenous time series. All data get Min-Max normalized between -1 and 1.

TABLE I. LIST OF EXOGENOUS TIME SERIES USED

Nr	Time series	Nr	Time series
1	Biomass	11	Hydro pumped storage generation
2	Hydro run-of-river and poundage	12	Hydro pumped storage consumption
3	Wind offshore	13	Other conventional
4	Wind onshore	14	Other forecasted
5	Photovoltaics	15	Grid load
6	Other renewable	16	White noise 1
7	Nuclear	17	White noise 2
8	Brown coal	18	Random walk 1
9	Hard coal	19	Random walk 2
10	Fossil gas		

A. *Feature Extraction*

1) Fourier Transformation: A fundamental function for signal analysis is the Discrete Fourier Transformation (DFT), which allows for the analysis of different frequencies in the time series. The presence of cycles in many elements, such as the working day and the deviation from the working rhythm at weekends, is a consequence of human habits. This principle also applies to the rhythm of other elements, including the position of the sun and the influence of the seasons over the course of the year. To a certain extent, these conditions are reflected in electricity generation and the price of electricity. The Fast Fourier Transformation (FFT) is an efficient algorithm that implements the DFT. One issue that can arise in spectral analysis is the leakage effect, which occurs due to the inaccurate capture of frequency resulting from the restricted discrete period under consideration and the finite number of discrete sampling points. There are several methods to mitigate this effect. We used a window function to weight the sampled values within the time series. Various window functions exist for this purpose. The Hanning window was selected due to its ability to effectively resolve frequencies and minimize leakage effects, making it well-suited for general data analysis [18], [19]. FFT is a highly effective means of identifying seasonality [20].

2) Wavelet Transformation: The FFT transforms the time series entirely from the time domain to the frequency domain, thereby causing the loss of direct temporal information. Consequently, after the transformation, only information of frequencies is available, not their occurrence in the time series.

This implies that no frequency changes can be detected within the time series. The WT is then employed to address this limitation. It combines the time and frequency domains, thereby enabling the detection of frequency changes within the time series. The short-time Fourier transform addresses this issue by employing a fixed window over the time series and transforming it accordingly. However, this approach limits the resolution of the frequencies, which is not the case with the WT [21]. The Daubechies type wavelet function of category four is used as the mother wavelet. This particular wavelet function was selected because of its optimal balance between time and frequency resolution [6], [22]; a decomposition depth of three is also selected. This choice enables a comparison to be conducted on four distinct levels: three levels of detail and the remaining information in the approximation. The feature selection methods are then employed on each level individually and subsequently aggregated. Mathematically, this is done for mRMR and partial correlation, and per majority decision for all others.

3) ARIX Prozess: The ARIX time series process represents a modification of the ARIMA process from the Box-Jenkins method [3]. ARIX is an acronym for Autoregressive (AR), Integration (I), and Exogenous (X). In comparison to the original process, ARIX incorporates exogenous data and omits the moving average (MA), as it is based on the calculation of the error of the previously calculated forecast values, which are not available during preprocessing. The structure and methods employed are rooted in [23],[24]. AR characterizes the relationship between actual and immediate past values, also termed lagged values. To address non-stationarity, which can be attributed to factors such as trends, seasonality, or structural changes, the I-part employs reference values that are shifted by τ according to the time-series. The weekday calendar method [23] is utilized to calculate τ. This method involves the estimation of a reference distance to past values for each day of the week (Monday to Sunday). This calculated distance τ then determines which data point from the past is used for Integration values. For instance, if $\tau_{\mathrm{Mon}} = 3$, the reference values for Mondays are from the previous Friday. In the context of time-series analysis, neural networks utilize mostly the same input data as described, but the key difference is that neural networks do not require a specific regression equation and are capable of learning non-linear relationships and latent features. By using neuronal networks additional exogenous variable like calendar features can be incorporated such as time of day, day of week, day of year, and holidays, in accordance with the calendar dependencies of the majority of power-related time series. These are calculated from the time stamp of each data point. Conversely, temporal features with periodic patterns can be represented through one-hot encodings, resulting in high-dimensional data input, or as sine and cosine transformed features with lower dimensions [25][26]. For every individual forecast timepoint, a tuple of data points used per ARIX process is calculated and then put together to a new higher dimensional time series.

B. Feature Selection

1) mRMR & partial correlation: Two common filter methods are minimum Redundancy Maximum Relevance (mRMR) and partial correlation (pcorr). The mRMR method involves selecting a set of all time series so that the average information content between the selected features and the target time series is maximized, while the information content between the time series in the set is minimized. For redundant features, only one should be selected. This approach requires a complete calculation of all combinations to find the optimal solution, rendering the use of forward or backward selection a common practice in determining the set.

The pcorr function can calculate the correlation between two variables, accounting for dependencies on other features. In this scenario, the aim is to calculate the correlation between individual time series and the target time series, with the aim of removing the influence of other time series. To achieve this, the partial correlation of a higher order must be calculated, which can be calculated recursively [27]. We selected the Pearson method, as it is more suitable for continuous data than the Spearman or Kendall methods [28].

2) LASSO & Elastic-Net: Least Absolute Shrinkage and Selection Operator (LASSO) [29] has shown to be well suited for feature selection due to the regression and regularization used. It is applied in various fields [30], including applications with time series [31]. LASSO is based on linear regression and minimizes the sum of the squared error between the calculated value and the target value. An upper limit is incorporated as a constraint via the sum of the absolute values of the model parameters, and the resulting optimization problem is equivalent to equation (4) [32].

$$\hat{\beta}(\lambda) = argmin_{\beta}\left(\frac{||Y-X\beta||_2^2}{T} + \lambda_1||\beta||_1\right) \tag{4}$$

In cases where multiple features are associated with each other through correlation, LASSO typically selects only one feature. However, when both features are to be selected, Elastic-Net [33] can be employed as a combination of LASSO and ridge regression. Elastic-Net employs both the L1 norm of LASSO and the L2 norm of ridge regression, with factor λ_2, as regularization mechanisms. The combination of λ_1 and λ_2 is dependent on each other, with α being defined as $\lambda_1/(\lambda_1 + \lambda_2)$. This allows for the adjustment of the strength between LASSO with $\alpha = 1$ and ridge regression with $\alpha = 0$. To ensure that the advantage of the corners in the search space is not lost, the range of possible values for the ratio $\alpha \in [0,1)$. The new optimization problem is as in (5).

$$\hat{\beta}(\lambda) = argmin_{\beta}\left(\frac{||Y-X\beta||_2^2}{T} + \alpha||\beta||_1 + (1-\alpha)||\beta||^2\right) \tag{5}$$

Cross-validation is used to determine hyperparameters for LASSO and Elastic-Net.

3) Dropout Layer: Dropout learning is an embedded feature selection method in which the prediction model learns directly which input data should be used. This is achieved by an additional layer, the dropout layer, situated between the input and the remaining network. This layer is capable of processing input data exclusively in binary form, so it can switch individual input data on or off. It is important to note that this does not refer to the commonly used dropout for individual neurons, which is used as an addition in network layers to prevent overfitting. The dropout layer method was developed independently in two variants, Dropout Feature Ranking (DFR) [12] and Stochastic Gates (SG) [13], each with a different approach. Both variants use a Bernoulli distribution to constrain the input data, but they differ in their relaxation function for the backpropagation of the error for the dropout layers. DFR utilizes a concrete distribution with values ranging from zero to one and with mass at both ends, while SG employs a relaxation based on Gaussian distributed Bernoulli variables to select features. DFR is used for feature selection, in addition to the relaxation previously outlined. It employs a modified loss function that considers the number of features utilized and aims to minimize them. Consequently, the function ensures that the error increases with the inclusion of more features, thereby achieving an additional filtering effect that eliminates features that contribute little or are redundant. For the feature selection, the layer is included in our A-FNN model, as it is the only neuronal network we used, that can be adapted to this layer.

C. Neuronal Networks

1) Linear regression model (LRM): LRM is a straightforward and efficient network that aims to estimate the desired output variable by employing a linear combination of the input variables. A notable drawback of this network is its inability to always map the interactions accurately between features and target, as these interactions may exceed the linear relationship.

2) Additive-Feedforward neuronal network (A-FNN): A-FNN is a multi-layer perceptron (MLP) Model with two input layers. Both inputs are processed by separate dense layers. These results are concatenated and processed further by the remaining network. The presence of two distinct initial layers enables the acceptance of two matrices of varying width and depth. This design enables the network to adapt to the preprocessing function of the ARIX process.

3) Time-series Dense Encoder (TiDE): TiDE is a new published prediction network for time series that utilizes an encoder-decoder approach [34]. In the encoder section, the dimensionality of the input variables is reduced by feature projection. Furthermore, a temporal decoder is employed for the output, into which the previous decoder and the features with future values flow. Consequently, these features possess a more direct path and a greater influence on the forecast. TiDE is designed for long-term forecasts, with the test commencing at 96 time steps. The exogenous variables used are limited due to the increased complexity and run time of the model. For this reason, we omitted nuclear energy from our dataset to increase performance, as this energy type in now longer used in Germany. The "other forecast" time series is always used, as it was a time series with high correlation in the model investigations. This selection reduces the brute force effort by a factor of 4.

V. EXPERIMENTS AND RESULTS

A. Evaluation of brute force neuronal networks

1) LRM: The result for LRM is displayed in TABLE II. , presenting a clear pattern within the top 10 feature combinations sorted by mean absolute error (MAE). Renewable energies (3-6), brown coal (8), and the other forecasted electricity sources (14) are consistently present, while nuclear energy (7), grid load (15), and the other conventional energy (13) are almost absent. This further shows the presence of features that exert a negative effect on the model and should be eliminated.

2) A-FNN: These results are displayed in TABLE III. Similar to the LRM, clear patterns emerge; however, distinctions are also apparent. Renewable energies (3-6) are all selected, while brown coal (8) and the other forecast electricity sources (14) are consistently present. Conversely, nuclear energy (7), grid load (15), and the other conventional energy sources (13) are almost absent, thereby showing a more significant difference to the LRM. The selection and configuration of a model have a profound impact on the time series used.

3) TiDE: The evaluation of the TiDE model in TABLE IV. is distinct as the MAE is shared by a total of 45 feature combinations. This model shows a greater degree of stability in the case of alterations to the feature set, showing an MAE difference of only 2.8% between the best and worst outcomes. Due to the high robustness of the model, the evaluation process is challenging, as altering the call parameters or the seed would likely result in changes to the order of the results. For the comparison with the feature selections, the combinations of the top 45 features are considered. In general, the TiDE model shows satisfactory performance, primarily due to its internal processing mechanisms, including feature projection and the encoder-decoder network. However, when evaluated against the optimal result achieved by LRM, TiDE exhibits a 30% reduction in performance. The number of encoder and decoder layers in TiDE, particularly the decoder output dimension, exerts a substantial influence on the result, in conjunction with the number of features that are fed into the model. Consequently, the brute force approach proves not to be optimal in generalizing the hyperparameters to the same value for all feature combinations. This approach is associated with a significantly higher effort and greater complexity in finding the corresponding parameters individually. Furthermore, although the forecast with 96 time steps falls within the possible application of the model optimized for long-term forecasting, it is only the lowest tested case.

B. Evaluation of FE and FS combinations

1) mRMR & partial correlation: As both methods are not able to differentiate between past and future values, only the time series with values from the past are used for the three renewable energies. The same applies to the use of ARIX preprocessing, as this selects reference values that are as close as possible to the current values. Results are shown in 0 We used a threshold of 0.005 for selection to filter out minor similarities due to random noise. The standard use of mRMR with plain normalization is not applicable in this case, as it selects almost all features, including noise. Raising the threshold to include only features with higher relevance and exclude the random walk, would also exclude important ones, such as 3 - 5 or 15. Photovoltaics (5) is classified as having low relevance, which is probably due to the many zero values in the time series, as no electricity is produced at night. Consequently, the information content is minimal during a significant portion of the time series, thereby strongly contradicting the high relevance observed during daylight hours. As a result, temporally limited influences are unable to be mapped effectively. Although the combination of WT yields better results compared to FFT all results do not match those of the brute force approach. For the Feature selection with partial correlation in TABLE VIII. , a limit is also set for the selection. Since it is not relevant if the value is positive or negative, but only the distance to 0, a minimum correlation of $\pm$ 0.1 is selected, thereby filtering out the disturbance variables. The selection with only the normalization fits the LRM model well, with a deviation of 2,2% from the optimum. The other two preprocessing methods yield results that are analogous to mRMR. In conclusion, it can be stated that partial correlation is significantly more effective than mRMR and corresponds more closely to the feature sets. The signal processing does introduce new values and a distinct perspective, which filters out errors but, purely mathematical processing is not effective and do not match with the selection of the brute force approach.

2) LASSO & Elastic-Net: As with mRMR and pcorr, both FS are limited in the same way. The combination with the ARIX preprocessing did not lead to any good evaluable calculations, which is due to the large amount of data. A threshold value of 0.005 is also selected for the evaluation as for mRMR. The evaluation of LASSO in TABLE IX. is initiated with pure normalization, resulting in the filtration of noise and a heightened selectivity compared to mRMR, yet still less than pcorr. The application of additional pre-processing does not yield favorable outcomes with FFT, and with WT solely the added noise time series are filtered out, while all others are retained. This combination of preprocessing and FS enables to be used as a preliminary step in another feature selections, by producing a cleaned feature set. For Elastic-Net in combination with normalization or FFT, a L1 ratio of 1 was calculated through cross-validation. This leads to identical results as obtained by LASSO. For the WT a strongly deviating value from the L1 ratio is calculated: 0.05. The result is displayed in TABLE X. As with LASSO, the additional noises are filtered out. Two other variables, including the undesired size 13, could be filtered out by changing the threshold, resulting in a superior selection compared to LASSO.

3) Dropout Layer: Due to the long training period of the model, most values converge to the boundary values (0 or 1) and the threshold can therefore be chosen roughly above the middle value of 0.5, we choose 0.6. The results of all preprocessing combinations, shown in TABLE XI. , do not give optimal results. The different resolutions may be a problem. The price signal has been upsampled to a 15-minute resolution to match all other time series. Alternatively, they could be downsampled and aggregated to hourly resolution. This procedure is examined in conjunction with the ARIX processing, results are listed in TABLE XII. Removing the noise from the input data has a stronger influence on the selection, as other time series are no longer selected. A similar effect can be observed when limiting to past values. The dropout MLP is therefore very sensitive to changes in the input data. In TABLE XIII. , the same observations are made with pure normalization. After removing noise and limiting to past values the selection is similar to that of LRM, but worse than that of pcorr. The ARIX preprocessing added the encoding of the calendar information to the brute force for the A-FNN, which can be considered as an additional input matrix. This information can also be important for feature selection. Therefore, this matrix is also added, but behind the drop-down layer, so that it is always included. The difference is clearly visible and the result fits very well with the A-FNN brute force selection.

VI. CONCLUSION AND OUTLOOK

We investigated which data is useful for day-ahead electricity price forecasting and how to select it by comparing different feature extraction and selection combinations against a brute force approach. The results of the brute force approach shows that the selection of features depends strongly on the model architecture, preprocessing, and additional data. This makes it difficult to create a uniform feature selection that can be generalized to all models. In future studies, models with similar architectures and their influence on the selected features should be examined. Furthermore, other use cases in the field of time series forecasting, especially in the energy sector (e.g. electricity or heat demand), should be considered to investigate a possible transferability of the results. Combinations of WT and LASSO or Elastic-Net have shown positive results for filtering out time series of pure noise. The differences in resolution between the target and exogenous time series poses a problem. As shown, the resolution of the exogenous should be adapted to that of the target time series. In June 2025, the day-ahead auction of Europe's energy markets will widely transition to a 15-minute Market Time Unit. As all time series will have the same resolution, the price forecasting and ability to select features should improve. This should be verified by upcoming studies. Another challenge is the choice of parameters. The results presented are always the best selection of the hyperparameters tried. For some combinations, changing the learning parameter affects the selection significantly. Therefore, an additional hyperparameter optimizer should be included in further studies. It has been shown that additional information given to the model should also be part of the feature selection. It is important to note exactly which data to use. The data from the ARIX process is too similar to the original data to be used in a feature selection. Other extracted values for date and time are relevant because they contain the necessary time dependency of the time series and should therefore be included in feature selections.

VII. REFERENCES

[1] J. Li *et al.,* "Feature Selection: A data perspective," *ACM Comput. Surv.*, vol. 50, no. 6, pp. 1–45, 2017, doi: 10.1145/3136625.

[2] H. Xiong, G. Pandey, M. Steinbach, and V. Kumar, "Enhancing data analysis with noise removal," *IEEE Trans. Knowl. Data Eng.*, vol. 18, no. 3, pp. 304–319, 2006, doi: 10.1109/TKDE.2006.46.

[3] G. E. P. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time series analysis: forecasting and control*: John Wiley & Sons, 2015.

[4] Rafał Weron, "Electricity price forecasting: A review of the state-of-the-art with a look into the future," *International Journal of Forecasting*, vol. 30, no. 4, pp. 1030–1081, 2014, doi: 10.1016/j.ijforecast.2014.08.008.

[5] Deepak Singhal and K.S. Swarup, "Electricity price forecasting using artificial neural networks," *International Journal of Electrical Power & Energy Systems*, vol. 33, no. 3, pp. 550–555, 2011, doi: 10.1016/j.ijepes.2010.12.009.

[6] A. J. Conejo, M. A. Plazas, R. Espinola, and A. B. Molina, "Day-ahead electricity price forecasting using the wavelet transform and ARIMA models," *IEEE Transactions on Power Systems*, vol. 20, no. 2, pp. 1035–1042, 2005, doi: 10.1109/TPWRS.2005.846054.

[7] P.-H. Kuo and C.-J. Huang, "An Electricity Price Forecasting Model by Hybrid Structured Deep Neural Networks," *Sustainability*, vol. 10, no. 4, 2018, doi: 10.3390/su10041280.

[8] Wendong Yang, Shaolong Sun, Yan Hao, and Shouyang Wang, "A novel machine learning-based electricity price forecasting model based on optimal model selection strategy," *Energy*, vol. 238, p. 121989, 2022, doi: 10.1016/j.energy.2021.121989.

[9] Kin G. Olivares, Cristian Challu, Grzegorz Marcjasz, Rafał Weron, and Artur Dubrawski, "Neural basis expansion analysis with exogenous variables: Forecasting electricity prices with NBEATSx," *International Journal of Forecasting*, vol. 39, no. 2, pp. 884–900, 2023, doi: 10.1016/j.ijforecast.2022.03.001.

[10] Salih Gunduz, Umut Ugurlu, and Ilkay Oksuz, "Transfer learning for electricity price forecasting," *Sustainable Energy, Grids and Networks*, vol. 34, p. 100996, 2023, doi: 10.1016/j.segan.2023.100996.

[11] Wei Li and Denis Mike Becker, "Day-ahead electricity price prediction applying hybrid models of LSTM-based deep learning methods and feature selection algorithms under consideration of market coupling," *Energy*, vol. 237, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0360544221017916

[12] Chun-Hao Chang, Ladislav Rampasek, and Anna Goldenberg, "Dropout Feature Ranking for Deep Learning Models," 2018.

[13] Y. Yamada, O. Lindenbaum, S. Negahban, and Y. Kluger, "Feature Selection using Stochastic Gates," in *Proceedings of the 37th International Conference on Machine Learning*, 2020, pp. 10648–10659. [Online]. Available: https://proceedings.mlr.press/v119/yamada20a.html

[14] Shahin Boluki, Randy Ardywibowo, Siamak Zamani Dadaneh, Mingyuan Zhou, and Xiaoning Qian, "Learnable Bernoulli Dropout for Bayesian Deep Learning," 2020. [Online]. Available: https://arxiv.org/abs/2002.05155

[15] Catalão, João Paulo da Silva, H. M. I. Pousinho, and V.\'\i. M. F. Mendes, "Hybrid wavelet-PSO-ANFIS approach for short-term electricity prices forecasting," *IEEE Transactions on Power Systems*, vol. 26, no. 1, pp. 137–144, 2010.

[16] P. Mandal, A. U. Haque, J. Meng, A. K. Srivastava, and R. Martinez, "A novel hybrid approach using wavelet, firefly algorithm, and fuzzy ARTMAP for day-ahead electricity price forecasting," *IEEE Transactions on Power Systems*, vol. 28, no. 2, pp. 1041–1051, 2013, doi: 10.1109/TPWRS.2012.2222452.

[17] Naoki Masuda, Mason A. Porter, and Renaud Lambiotte, "Random walks and diffusion on networks," *Physics Reports*, 716-717, pp. 1–58, 2017, doi: 10.1016/j.physrep.2017.07.007.

[18] Pico Technology, "Fast Fourier Transforms explained," Accessed: Mar. 18 2025. [Online]. Available: https://www.picotech.com/download/brochures/PT-012746-DO-2-AR501_Fast_Fourier_Transform_%28FFT%29_white_paper.pdf

[19] Emerson - NI, *Grundlegendes zu FFTs und Fensterfunktionen.* [Online]. Available: https://www.ni.com/de/shop/data-acquisition/measurement-fundamentals-main-page/analog-fundamentals/understanding-ffts-and-windowing.html (accessed: Mar. 18 2025).

[20] H. Musbah, M. El-Hawary, and H. Aly, "Identifying Seasonality in Time Series by Applying Fast Fourier Transform," in *2019 IEEE Electrical Power and Energy Conference (EPEC)*, 2019, pp. 1–4.

[21] L. Chun-Lin, "A tutorial of the wavelet transform," *NTUEE, Taiwan*, vol. 21, p. 22, 2010.

[22] N. Amjady and F. Keynia, "Short-term load forecasting of power systems by combination of wavelet transform and neuro-evolutionary algorithm," *Energy*, vol. 34, no. 1, pp. 46–57, 2009, doi: 10.1016/j.energy.2008.09.020.

[23] S. Klaiber, "Analyse, Identifikation und Prognose preisbeeinflusster elektrischer Lastzeitreihen," Ilmenau, 2020. [Online]. Available: https://www.db-thueringen.de/receive/dbt_mods_00046851

[24] L. Richter, T. Bender, S. Lenk, and P. Bretschneider, "Generating Synthetic Electricity Load Time Series at District Scale Using Probabilistic Forecasts," *Energies*, vol. 17, no. 7, 2024, doi: 10.3390/en17071634.

[25] M. G. Pinheiro, S. C. Madeira, and A. P. Francisco, "Short-term electricity load forecasting—A systematic approach from system level to secondary substations," *Applied Energy*, vol. 332, p. 120493, 2023, doi: 10.1016/j.apenergy.2022.120493.

[26] A. Gasparin, S. Lukovic, and C. Alippi, "Deep learning for time series forecasting: The electric load case," *CAAI Trans on Intel Tech*, vol. 7, no. 1, pp. 1–25, 2022, doi: 10.1049/cit2.12060.

[27] M. Lange, "Partielle Korrelationen und Partial Mutual Information zur Analyse von fMRT-Zeitreihen," Hochschule Mittweida, 2011.

[28] N. S. Chok, "Pearson's versus Spearman's and Kendall's correlation coefficients for continuous data," University of Pittsburgh, 2010.

[29] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society Series B: Statistical Methodology*, vol. 58, no. 1, pp. 267–288, 1996.

[30] R. Muthukrishnan and R. Rohini, "LASSO: A feature selection technique in predictive modeling for machine learning," in *2016 IEEE International Conference on Advances in Computer Applications (ICACA)*, 2016, pp. 18–20.

[31] Jiahan Li and Weiye Chen, "Forecasting macroeconomic time series: LASSO-based approaches and their forecast combinations with dynamic factor models," *International Journal of Forecasting*, vol. 30, no. 4, pp. 996–1015, 2014, doi: 10.1016/j.ijforecast.2014.03.016.

[32] V. Fonti and E. Belitser, "Feature selection using lasso," *VU Amsterdam research paper in business analytics*, vol. 30, pp. 1–25, 2017.

[33] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *Journal of the Royal Statistical Society Series B: Statistical Methodology*, vol. 67, no. 2, pp. 301–320, 2005.

[34] Abhimanyu Das, Weihao Kong, Andrew Leach, Shaan Mathur, Rajat Sen, and Rose Yu, "Long-term Forecasting with TiDE: Time-series Dense Encoder," 2023. [Online]. Available: https://arxiv.org/abs/2304.08424

VIII. APPENDIX

The following tables show the results of the experiments, with numbers referring to the features listed in Table I. Tables II through IV show the results of the three individual brute force approaches for the three networks. Green colored cells in the Top 10 indicate that this feature was included in this combination. The Top 10% row shows how often each feature was included in the combination. Table V combines the results of Tables II, III and IV, with green cells meaning the feature should be included and red meaning it should be excluded. Cells with no color are neutral features, that don't have a strong impact on the network whether they are included or not.

TABLE II. SELECTED FEATURES WITH LRM BASED ON THE BEST NORMALIZED MAE ACHIEVED

Top	MAE	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	0,059788				🟩	🟩					🟩	🟩	🟩	🟩		🟩
2	0,059912				🟩	🟩		🟩			🟩	🟩	🟩	🟩		🟩
3	0,060000				🟩	🟩					🟩		🟩	🟩		🟩
4	0,060062				🟩	🟩					🟩		🟩			🟩
5	0,060106				🟩	🟩		🟩			🟩		🟩	🟩		🟩
6	0,060106	🟩	🟩		🟩	🟩				🟩	🟩	🟩	🟩	🟩		🟩
7	0,060122	🟩			🟩	🟩					🟩	🟩	🟩	🟩		🟩
8	0,060125				🟩	🟩					🟩	🟩	🟩			🟩
9	0,060127	🟩	🟩		🟩	🟩				🟩	🟩		🟩	🟩		🟩
10	0,060134	🟩			🟩	🟩				🟩	🟩	🟩	🟩			🟩
10%	Occurrence of the feature [%]	46,7	51,7	28,4	58,2	99,5	38,7	44,6	25,3	55,6	60,7	47,7	100,0	47,2	59,9	65,4

TABLE III. SELECTED FEATURES WITH A-FNN BASED ON THE BEST NORMALIZED MAE ACHIEVED

Top	MAE	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	0,04615		🟩	🟩	🟩	🟩	🟩		🟩		🟩	🟩	🟩		🟩	
2	0,04628	🟩	🟩	🟩	🟩	🟩	🟩		🟩		🟩	🟩	🟩		🟩	
3	0,04647	🟩	🟩	🟩	🟩	🟩	🟩		🟩			🟩	🟩		🟩	
4	0,04647	🟩	🟩	🟩	🟩	🟩	🟩		🟩		🟩	🟩	🟩	🟩	🟩	
5	0,04648		🟩	🟩	🟩	🟩	🟩		🟩	🟩		🟩	🟩		🟩	
6	0,04650			🟩	🟩	🟩	🟩		🟩	🟩			🟩		🟩	
7	0,04651	🟩	🟩	🟩	🟩	🟩	🟩		🟩		🟩	🟩			🟩	
8	0,04658	🟩	🟩	🟩	🟩	🟩	🟩		🟩	🟩	🟩	🟩	🟩		🟩	
9	0,04658		🟩	🟩	🟩	🟩	🟩		🟩			🟩	🟩		🟩	
10	0,04663			🟩	🟩	🟩	🟩		🟩		🟩	🟩	🟩		🟩	
10%	Occurrence of the feature [%]	45,9	56,6	67,6	99,2	74,0	75,1	34,0	68,5	41,5	51,9	54,3	47,0	45,8	90,8	45,0

TABLE IV. SELECTED FEATURES WITH TIDE BASED ON THE BEST NORMALIZED MAE ACHIEVED

Top	MAE	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1-45	0,08586	28	27	45	45	45	26		30	28	20	23	33	2		26
10%	Occurrence of the feature [%]	60,7	60,8	50,6	50,7	50,4	60,4		60,8	60,2	60,9	60,9	60,8	60,4		60,6

TABLE V. SUMMARY OF SELECTED FEATURES OF NEURONAL NETWORKS

WModel	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
LRM			🟥	🟩	🟩	🟥		🟥		🟩		🟩			🟩
A-FNN			🟩	🟩	🟩	🟩	🟥	🟩						🟩	🟥
TiDE			🟩	🟩	🟩							🟩	🟥		

Tables VII through XIII show the results of the combinations of FE and FS. The Tables are for individual FS with the FE as rows. The cells contain the result value calculated by the respective method, with green cells meaning the feature of that column would be selected. If the cell is yellow, the feature is near the threshold and could be excluded without excluding other selected (green) features by changing the threshold. Not colored cells mean that the feature is not selected and if the cell is grey the feature was not used in this combination of FE and FS. Table VI shows the information and color coding described.

In the Tables XI to XIII columns 3,4 and 5 are split, because for those features both time series past (left) and forecasted future values (right) are analyzed separately.

Rows in Tables XII and XIII may have a "+", indicating that the specified change is in addition to those above. For example, in row three of Table XII, only past values are used in addition to the exclusion of noise features from row two.

TABLE VI. CELL COLORS AND MEANING FOR EVALUATION OF FE AND FS COMBINATION (TABLE VII TO XIII)

🟩 Feature selected	🟨 Feature selected, but near threshold	Feature not selected	▨ Feature not included

TABLE VII. SELECTED FEATURES FROM CALCULATED VALUES OF MRMR

mRMR	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
Norm.	🟩 0,061	-0,033	🟩 0,239	🟩 0,297	🟩 0,172	🟩 1,425	🟩 0,634	🟩 0,618	🟩 0,800	🟩 0,630	🟩 0,270	🟩 0,384	🟩 0,285	🟩 7,285	🟩 0,217	0,003	-0,002	🟩 0,335	🟩 0,431
FFT	🟩 0,014	🟩 0,028	-0,021	🟩 0,005	-0,006	-0,009	-0,007	-0,002	-0,003	0,003	🟩 0,014	🟩 0,008	-0,013	🟩 0,129	🟩 0,041	-0,018	-0,013	-0,026	-0,011
Wavelet	-0,035	🟩 0,015	-0,020	🟩 0,049	🟩 0,054	-0,176	-0,036	🟩 0,284	🟩 0,360	🟩 0,319	🟩 0,327	🟩 0,223	-0,157	🟩 1,310	🟩 0,371	-0,048	-0,039	-0,179	-0,235

TABLE VIII. SELECTED FEATURES FROM CALCULATED VALUES OF PARTIAL CORRELATION

pcorr	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
Norm.	-0,090	0,016	-0,043	🟩 -0,257	🟩 -0,221	-0,079	0,010	0,012	-0,073	-0,043	0,077	🟩 -0,106	0,083	🟩 0,193	🟩 0,173	0,000	-0,006	0,063	0,093
FFT	0,005	-0,031	🟩 0,200	0,045	0,074	-0,076	🟩 0,197	0,013	0,038	0,092	🟩 0,184	🟩 0,200	0,093	🟩 0,289	🟩 -0,196	0,009	0,005	-0,083	-0,036
Wavelet	0,096	0,084	0,063	🟩 0,250	🟩 0,235	🟩 0,105	🟩 0,133	0,073	0,053	0,047	🟩 0,161	🟩 0,173	0,098	🟩 0,753	🟩 0,319	0,016	0,036	0,065	0,095

TABLE IX. SELECTED FEATURES FROM CALCULATED VALUES OF LASSO

LASSO	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
Norm	🟩 -0,009	🟩 0,012	-0,004	🟩 -0,090	🟩 -0,048	🟩 -0,007	0,000	🟩 0,028	0,000	0,000	🟩 0,037	🟩 -0,040	0,000	🟩 0,173	🟩 0,063	0,000	0,000	0,001	0,002
FFT	0,000	🟩 -0,018	🟩 0,090	🟩 0,028	🟩 0,073	🟩 -0,039	🟩 0,180	🟩 0,014	🟩 0,047	🟩 0,128	🟩 0,199	🟩 0,267	🟩 0,063	🟩 0,631	🟩 -0,251	0,000	0,000	🟩 -0,095	🟩 -0,032
Wavelet	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟨	🟩	🟩	🟩	🟩	🟩	🟩				

TABLE X. SELECTED FEATURES FROM CALCULATED VALUES OF ELASTIC-NET

Elastic-Net	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
Norm	🟩 -0,009	🟩 0,012	-0,004	🟩 -0,090	🟩 -0,048	🟩 -0,007	0,000	🟩 0,028	0,000	0,000	🟩 0,037	🟩 -0,040	0,000	🟩 0,173	🟩 0,063	0,000	0,000	0,001	0,002
FFT	0,000	🟩 -0,018	🟩 0,090	🟩 0,028	🟩 0,073	🟩 -0,039	🟩 0,180	🟩 0,014	🟩 0,047	🟩 0,128	🟩 0,199	🟩 0,267	🟩 0,063	🟩 0,631	🟩 -0,251	0,000	0,000	🟩 -0,095	🟩 -0,032
Wavelet	🟨	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟨	🟩	🟩				

TABLE XI. SELECTED FEATURES OF DROPOUT WITH ORIGINAL FEATURE EXTRACTION

Dropout	1	2	3		4		5		6	7	8	9	10	11	12	13	14	15	16		17	18	19
Norm	🟩		🟩	🟩		🟩	🟩		🟩		🟩			🟩	🟩	🟩	🟩	🟩					🟩
ARIX	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩				🟩	🟩
FFT				🟩			🟩				🟩	🟩	🟩	🟩	🟩		🟩						
Wavelet			🟩	🟩		🟩	🟩							🟩	🟩		🟩	🟩					

TABLE XII. SELECTED FEATURES OF DROPOUT WITH ARIX PREPROCESSING AND ADDITIONAL MODIFICATIONS

Dropout	1	2	3		4		5		6	7	8	9	10	11	12	13	14	15	16	17	18	19
All features	🟩	🟩	🟩	🟩		🟩	🟩	🟩		🟩			🟩			🟩	🟩	🟩				🟩
With out noise	🟨		🟩	🟩		🟩	🟩	🟩	🟩	🟩	🟩					🟩	🟩	🟩	▨	▨	▨	▨
+ Only past time series	🟩	🟨	🟩							🟩	🟩	🟩				🟩	🟩	🟩	▨	▨	▨	▨

TABLE XIII. SELECTED FEATURES OF DROPOUT WITH NORMALIZATION AND ADDITIONAL MODIFICATIONS

Dropout	1	2	3		4		5		6	7	8	9	10	11	12	13	14	15	16	17	18	19
All features	🟩	🟩		🟩		🟩		🟩			🟩			🟩	🟩		🟩	🟩			🟩	🟩
With out noise		🟩		🟨		🟩		🟨			🟩			🟩	🟩		🟩	🟩	▨	▨	▨	▨
+ Only past time series		🟩		▨	🟩	▨	🟩	▨						🟩	🟩		🟩	🟩	▨	▨	▨	▨
+ calendar features		🟩		▨	🟩	▨	🟩	▨	🟩		🟩			🟩	🟩		🟩		▨	▨	▨	▨