Author Accepted Manuscript version of the paper by Stian Backe et al.
in 2023 19th International Conference on the European Energy Market - EEM (2023) DOI: http://dx.doi.org/10.1109/EEM58374.2023.10161961.

Predictions of prices and volumes in the Nordic balancing markets for electricity

Stian Backe
Energy Systems
SINTEF Energy Research
Trondheim, Norway
stian.backe@sintef.no

Signe Riemer-Sørensen
Mathematics and Cybernetics
SINTEF Digital
Oslo, Norway
signe.riemer-sorensen@sintef.no

David A. Bordvik
Hydropower
Akershus Energi
Oslo, Norway
david.andreas.bordvik@akershusenergi.no

Shweta Tiwari
Energy Systems
SINTEF Energy Research
Trondheim, Norway
shweta.tiwari@sintef.no

Christian Andre Andresen
Energy Systems
SINTEF Energy Research
Trondheim, Norway
christian.andresen@sintef.no

***Abstract*—The electricity market is driven by complicated interactions that are hard to model analytically. This is particularly the case for the balancing market, where imbalances between supply and demand after the day-ahead market clearance are balanced. The balancing market bridges the gap between the day-ahead market and the actual power system operations. Being able to predict the necessary balancing volumes and prices some hours in advance of the operational hour will allow power producers to plan their production and trading in a more optimal way. There exist large amounts of open data that could contain predictive information about the balancing market, including day-ahead market data and climatic data. However, the literature on forecasting volume and prices in the balancing market is sparse compared to the rich literature on forecasting for the day-ahead market. Neural networks are powerful functional approximators and well-suited to model the complex relationships in the power market. It may also be used to study the predictability of the balancing volumes and prices forward in time. In this paper, we develop a model based on long short-term memory (LSTM) recurrent neural networks to predict volumes and prices in the Nordic balancing market based on public accessible data. Results show that the LSTM model performs well when compared to the two baselines selected. However, the performance is not significantly better, which indicates that the market data does not hold significant predictive information.**

***Index Terms*—power generation planning, balancing market, machine learning, recurrent neural networks, forecasting**

I. Introduction

In a deregulated power market, a large part of the electricity volumes to be produced are sold one day ahead in the day-ahead market, commonly referred to as the spot market. However, there are many factors that can change both the realised consumption and production after the day-ahead market has been settled. Origins of the changes can be related to weather, day-ahead production volumes and prices, capacities in the transmission grid, and human actors in the market. This introduces the need for balancing services and markets. For example, if the actual wind speed is less than forecasted in a region with wind power generation, the resulting power production will be smaller than planned for. Similarly, lower temperatures will increase consumption for heating. To meet their commitments in the day-ahead market and adjust their position, the producers can either trade their imbalance through bilateral agreements in the intraday market (such as SIDC[1]) or settle them in the balancing power market [18].

The ability to predict the necessary balancing volumes some hours ahead of the operational hour will allow for improved production planning and support the strategic decisions of whether to settle imbalances in one market or the other. The producers have access to a vast array of data sources for this planning. Some are public such as the available transfer capacity (set by the transmission system operator (TSO)), urgent market messages (UMMs), some market data (e.g., ENTSO-E Transparency Platform) and public weather forecasts. Additional data sources are either proprietary to the specific producers (plant details or internal water values), while others can be acquired from vendors (specialised weather services, consultancy supplied water values, or market predictions). In addition to this, the TSO has access to proprietary data (grid state, reserve availability) that can not be accessed by the producers. See the IMPALA project for an application[2]

The complexity of the system and a relatively large amount of historical data motivates a deep learning modelling approach. Neural networks are very powerful functional approximators that can be trained to generalise over complex relationships in the data. The balancing volumes and prices have a temporal dependency, and their values depend on activity in the day-ahead market. Thus, we propose a model structure of recurrent neural networks with multivariate time series. This structure has been proven to perform well in several time series prediction cases [12, 13].

[1] https://www.epexspot.com/
[2] https://www.statnett.no/contentassets/511d2161889b4ed9affdedd14aadabf8/2019-impala.pdf

We compare related works and highlight the need to explore imbalance predictions using open data in Sec. II. In Sec. III, we introduce the structure of the recurrent neural network model (III-A) and baseline models and comparison metrics (III-B). The results are presented and discussed in Sec. V and Sec. VI concludes the paper.

II. Related works and our contribution

The literature on balancing volumes and imbalance prices in the Norwegian and Nordic power markets is scarce [8, 10, 11, 18]. Klæboe et al. [10] found that coordinated bidding between the day-ahead and balancing market is currently not profitable, but it might become so with increased volumes and imbalance price premiums, i.e., differences between day-ahead price and imbalance price. However, their study focused on imbalance prices and not on predicting balancing volumes for production planning. Research on balancing volumes is also scarce for regions outside Norway. Since the balancing is strongly dependent on local conditions and definitions of market zones, only the methods would be transferable, not the results themselves.

For the day-ahead and intraday markets, the literature is more abundant with individual studies [5, 6, 7, 11, 12] and reviews of commonly used methods and benchmarks [13]. Kuo and Huang [12] provide an example of deep learning for price prediction in the North Eastern USA day-ahead market. The results are extensively compared with other known methods, including support vector machines, random forests, multi layer perceptron, and pure convolutional neural networks (CNN) and long short-term memory (LSTM) approaches. Kuo and Huang [12] concludes that the combination of LSTM and CNN is superior.

Klæboe et al. [11] covers both balancing volume and imbalance price forecasting in the Norwegian balancing market, but mainly focus on statistical time series-based forecasting models, including autoregressive models, Markov models, and arrival rate models for predicting the balancing state. The study considered market data from 2010-2012, and conclude that current design of the balancing market as a means to handle unforeseen events leads to randomly distributed balancing volumes and imbalance prices.

Salem et al. [17] and Bottieau et al. [3] also addresses forecasting of both balancing volumes and imbalance price premiums, but from the TSO's perspective based on non-public historical imbalance data. Salem et al. [17] successfully apply a quantile regression forest (QRF) approach. This is an ensemble learning method with the ability to generate prediction intervals. The methodology was trained and evaluated on imbalance data from Norway for 2015-2016 using two hours prediction horizon. Bottieau et al. [3] presents a quantile regression tool producing a probabilistic prediction of the future system imbalance. Furthermore, the tool quantifies risks and optimises participation of a market player in the imbalance market. The tool utilises an encoder-decoder neural network approach and is trained on historic imbalances along with historic production split on different sources. Recent developments of the electricity market and the existing studies [3, 12, 17] motivates our use of machine learning for forecasting of balancing volumes and imbalance price premiums from public data.

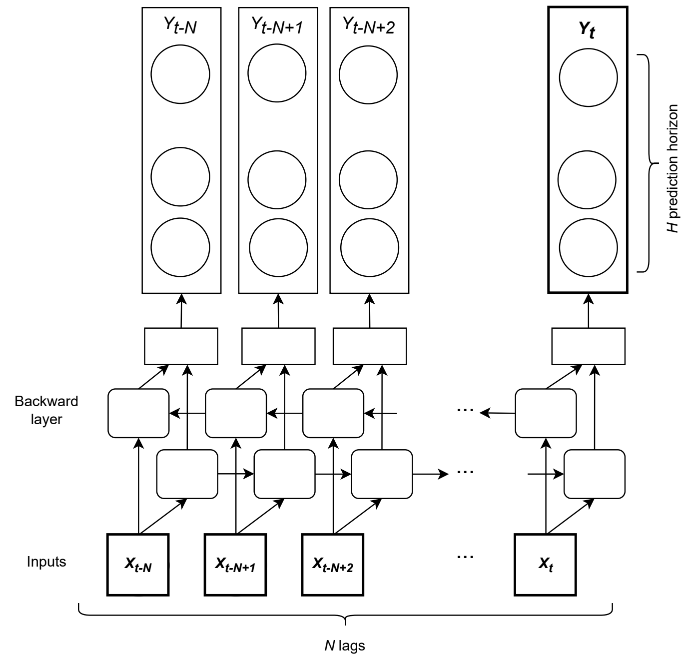

Fig. 1. Illustration of the bidirectional LSTM model.

There are also commercial actors that have developed their own methodology and are offering prediction tools and services which are partly based on data-driven methods. One such example is Optimeering [15] who offers operational forecasting services for the power sector. For the regulation market they offer a classification model predicting up or down regulation for a few hours ahead. The public insight into these commercial methodologies is limited.

To our knowledge, there are no published works applying the concept of recurrent neural networks to the prediction of volumes and prices in the balancing power markets in the Nordics using public data. This is thus a novel contribution to the field and differs from the approaches available so far. Based on results by Bordvik [2] in his master's thesis, we further explore an LSTM model since it outperforms Multiple Linear Regression, Support Vector Regression, Ridge Regression, and XGBOOST (see Figure 5.8 in [2]).

III. Methodology

A. The LSTM model

We treat the problem of predicting balancing volumes and prices given open market information as a supervised regression problem. We consider a bidirectional LSTM model for which the general architecture is presented in Fig. 1.

The model is given vectors $X_{t-N}, ..., X_t$ with N lags for one or several time series as input to bidirectional LSTM layers. The last known value(s) for the time series is X_t for time step t. The bidirectional structure of the LSTM layers means the model is trained to generalise past information from one or several time series with N lags in both chronological and reversed chronological order. For each time step, its respective bidirectional LSTM layer connects to dense layers with S neurons. These dense layers further connect to smaller dense layers of $\frac{S}{5}$ neurons before they connect to the target vectors Y. Note that for each time step, the corresponding dense layers are not connected to the dense layers of other time steps. We do not consider dropouts in part due to no over-fitting issues. The target vectors Y are the continued time series of volumes or prices for H future time steps for each time step, so the model is trained on $H \times N$ target values. The architecture allows for multiple output variables, but for simplicity we consider one model for each output.

B. Baseline models and comparison metrics

We are not aware of any similar models in the literature that can freely be used as a baseline model (see Sec. II). Hence we compare to two simple baselines. The first baseline (*Zero*) is to predict zero for all hours, which is close to the historic mean value for both balancing volume and imbalance price premium for all price zones. The second baseline (*Last*) is assuming all next hours will have the same value as the last known value. The models are compared with root mean squared error (RMSE) per look-ahead hour for each price zone individually.

Bordvik [2] also explore other baselines, such as multiple linear regression. However, it is found to perform similarly to *Last* and is therefore omitted in this paper.

IV. Training data and setup

A. Market data

For the demonstration, we focus on the Nordic power market and include the Norwegian price zones (NO1-5) plus their bordering market zones (SE1-3, DK1, and FI) for which regulation data is available at an hourly resolution from 2016 to 2023. We train and test the model on public historical data from Nord Pool downloaded via the Nord Pool Market Data API[3].

We train models with two different targets: Balancing volume and imbalance price premium. For balancing volume, we generate the target by summing the two time series for 'OrdinaryUpVolume' and 'OrdinaryDownVolume' from the Nord Pool API for every hour. For the few ($< 0.02\%$) missing values of balancing volume in the training and test data, we forward fill the last known value. For imbalance price, we generate the target as the difference between the time series for 'ConsumptionImbalancePrice'[4] and 'AreaPrice', effectively producing the imbalance price premium. All training data is scaled using MinMaxScalar from scikit-learn [16].

We train univariate LSTM models for each prize zone and for each of the two targets. Additionally, we train multivariate LSTM models for each zone and target including the following market feature time series as additional input:

- *RegBidUp*: The hourly volume of up regulating bids for the same zone ('TotalUpVolume')
- *RegBidDown*: The hourly volume of down regulating bids for the same zone ('TotalDownVolume')
- *Cons*: Settled consumption for the same zone ('TotalConsumption')
- *Prod*: Settled power production for the same zone ('TotalProduction')
- *DA price*: Day ahead clearing price for the same zone ('AreaPrices')
- *Wind*: The sum of settled wind production for all zones ('WindPower')

B. Training

The model is trained on data from January 2016 to December 2021 (last 20% used for validation during training). The model is tested and evaluated on unseen data from January 2022 to December 2022. Although the model always produces $H \times N = 100$ predictions when provided input data, we only evaluate the last $H = 10$ predictions as these predictions would be the ones used in practice. After imputation, the training and validation data consist of 52 590 hours of data and the test data of 8 751 hours.

The general neural network model is implemented in Python 3.9.7 using Keras 2.4.3 [4] and TensorFlow 2.4.1 [1]. The model is trained with mean squared error as the loss function using the Adam optimizer [9] with a learning rate of 0.005. Each neural network is trained for 100 epochs. We reduce the learning rate if the validation loss does not improve after 3 consecutive epochs. We train in batches of 5 samples per iteration, which yields 842 iterations per epoch. We have experimented with various hyperparameters and train models consisting of $N = 10$ recurrent lags, $S = 300$ neurons in the first dense layer, and $H = 10$ future time steps. The training and validation losses are visually inspected for convergence and overfitting[5].

V. Results and discussion

A. Comparing with additional features

Fig. 2 shows the distribution of the non-normalised RMSE for each additional feature, look-ahead hour, and each target. The univariate LSTM (*None*) generally outperforms all the multivariate LSTMs for the first look-ahead hour for both targets. However, for future look-ahead hours, *Wind*, *Cons*, and *Prod* provide minor improvements to the quality of the forecast for the imbalance premium target. The worst performing LSTMs for both targets are the models including day-ahead price as an additional feature. This could be due to day-ahead prices having a historic evolution that is significantly different from the two targets, which adds significant noise.

[3]https://marketdata.nordpoolgroup.com/

[4]The price for balancing in the dominating direction for each hour (either 'RegulatingUpPrices' or 'RegulatingDownPrices').

[5]When trained on a machine with 2x Xeon Gold 6126 (12 Cores, 2.4 GHz, HT) CPUs and 512 Gb RAM memory, each training takes approximately 20 s per epoch. Computation time for all epochs, targets, zones, and features is approximately $20 \times 100 \times 2 \times 10 \times 7 = 280\ 000$ s (78 hours).

Author Accepted Manuscript version of the paper by Stian Backe et al.
in 2023 19th International Conference on the European Energy Market - EEM (2023) DOI: http://dx.doi.org/10.1109/EEM58374.2023.10161961.

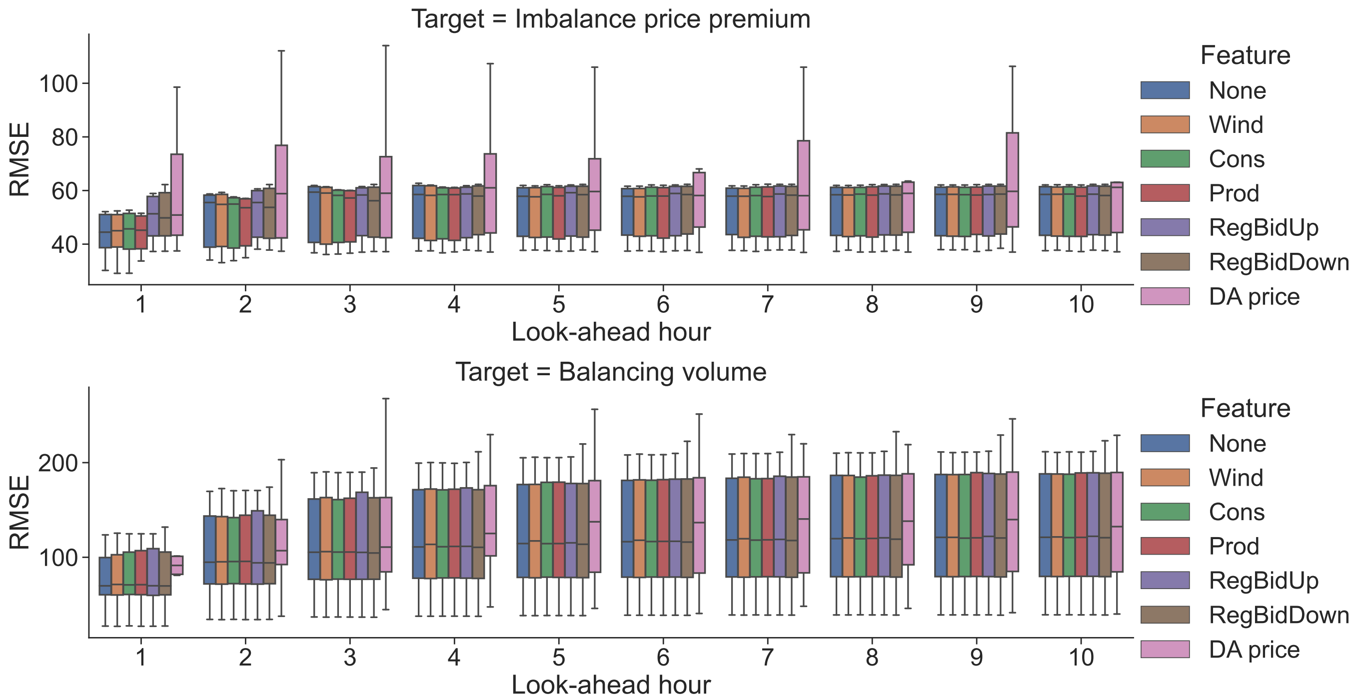

Fig. 2. Box plot of root mean squared errors (RMSE) on the test data for each feature, look-ahead hour, and target (without outliers).

B. Comparing with baselines

Fig. 3 shows the distribution of the non-normalised RMSE for the univariate LSTMs and the two baselines (*Zero* and *Last*) for each look-ahead hour. Compared to the baselines, the LSTM performs better when predicting balancing volume than predicting imbalance price premium. For the first look-ahead hour, the LSTM performs similar to or worse than *Last*. For future look-ahead hours, the LSTM performs similar to or better than *Zero*.

C. Comparing the price zones

Fig. 4 shows the non-normalised RMSE for the univariate LSTM and the two baselines (*Zero* and *Last*) for each price zone. For all price zones, the LSTM provides the overall best predictions when considering all look-ahead hours, followed by *Zero* and *Last*, but for some zones, the difference is marginal when predicting price premiums.

VI. Conclusion and outlook

Overall, the LSTM model for balancing volumes performs better than predicting zero or filling forward the last known value with target. For price premiums, the LSTM model struggles to outperform the baselines with target imbalance price premium. For both targets, the LSTM essentially learns to predict the last known value with minor adjustments for the first look-ahead hours. For later look-ahead hours, it learns to predict close to zero.

We have tried to train LSTMs towards targets equal to the hourly change in balancing volumes and price premiums. The resulting target series has more zero values, and the resulting LSTMs perform more similarly to *Zero* for all hours.

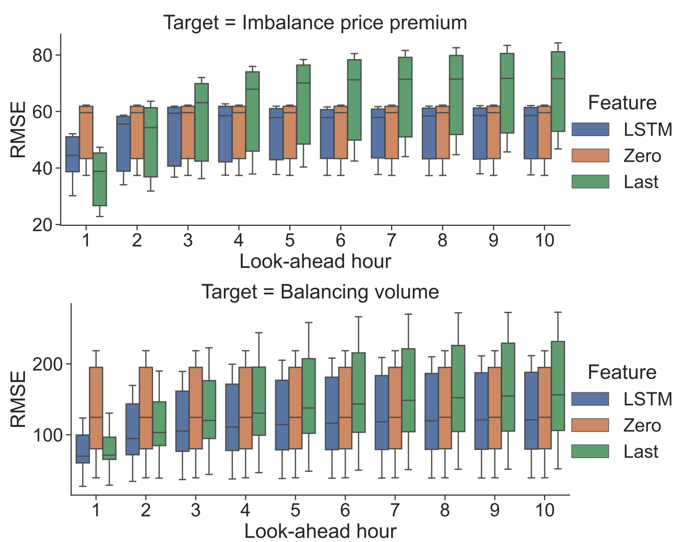

Fig. 3. Box plot of root mean squared errors (RMSE) on the test data for the univariate LSTM and the baselines for each look-ahead hour and target (without outliers).

The large values of RMSE indicates that the LSTM model is not a perfect forecasting strategy. We see two potential reasons for this performance of the LSTM:

1) The model does not get sufficient relevant information to predict future regulation. If the information content of the input features are not relevant, they will basically

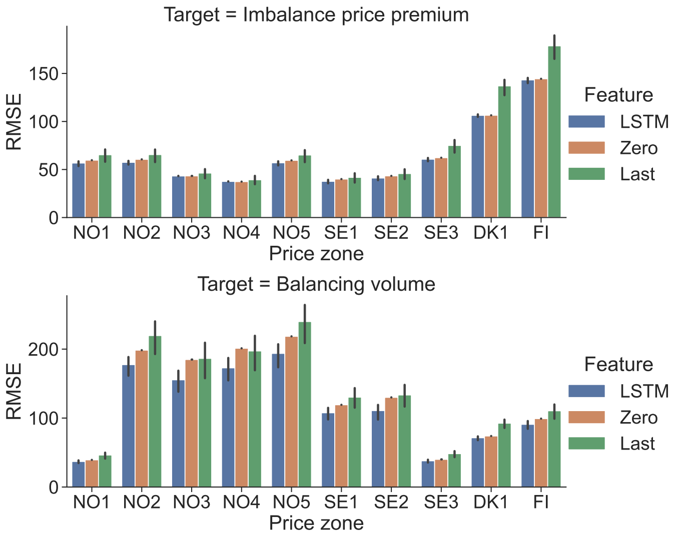

Fig. 4. Bar plot of root mean squared errors (RMSE) on the test data for each zone, feature, and target.

become noise for the model and have an adverse effect on the predictions.

2) The loss function used during training quantifies mean squared error, and predicting the last known regulation tends to produce relatively small mean squared error due to the presence of auto-correlation in the data (balancing events often happens over several subsequent hours).

The balancing has some auto-correlation, but this alone does not provide enough information to predict the future balancing with high quality. The challenge most likely originates in the fact that balancing depends on many other factors than the features tested here. Adding more features will provide more information for the model. However, the included features must be carefully selected such that they do not lead to larger increase in the noise than in information.

We know that electricity demand is weather dependent, so weather information is relevant to explore. Bordvik [2] investigate the predictive information of weather data based on the difference between the weather forecast available during day-ahead bidding and the latest available weather forecast. Bordvik [2] does not find that weather information improves the volume and price forecasting; however, there are numerous ways of including information about the weather, and continued exploration is relevant.

The challenge with the loss function is two-fold. Firstly, due to auto-correlation in the data, the loss does not grow very large when the prediction is very similar to the previous hour. Secondly, zero regulation is dominating the data, so the model is trying to learn a minority effect. This could potentially be overcome by adding a binary cross entropy term to the loss function to distinguish between regulation and non-regulation, and then combined with the mean squared error for increased quality on the predictions. Such approaches have been applied in image and video analysis but not for the power market [14].

Acknowledgment

This work was funded by the Research Council of Norway via the Energy-X grant number 309315 and the project partners Skagerak Kraft AS, Eviny Fornybar AS, and Equinor Energy AS.

References

[1] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, and *et al.* TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. URL https://tensorflow.org. Software available from tensorflow.org.

[2] D. A. Bordvik. Forecasting regulation market balancing volumes from market data and weather data using deep learning and transfer learning. Master's thesis, 2022.

[3] J. Bottieau, L. Hubert, Z. D. Grève, F. Vallée, and J.-F. Toubeau. Very-Short-Term Probabilistic Forecasting for a Risk-Aware Participation in the Single Price Imbalance Settlement. *IEEE Transactions on Power Systems*, 35(2): 1218–1230, 2020. doi: 10.1109/TPWRS.2019.2940756.

[4] F. Chollet et al. Keras. https://keras.io, 2015.

[5] I. Dimoulkas, M. Amelin, and M. R. Hesamzadeh. Forecasting balancing market prices using Hidden Markov Models. In *2016 13th International Conference on the European Energy Market (EEM)*, pages 1–5, June 2016. ISSN: 2165-4093.

[6] Z. Hameed, S. Hashemi, and C. Træholt. Applications of AI-Based Forecasts in Renewable Based Electricity Balancing Markets. In *2021 22nd IEEE International Conference on Industrial Technology (ICIT)*, volume 1, pages 579–584, Mar. 2021.

[7] L. Hirth and I. Ziegenhagen. Balancing power and variable renewables: Three links. *Renewable and Sustainable Energy Reviews*, 50:1035–1051, Oct. 2015. ISSN 1364-0321.

[8] S. Jaehnert, H. Farahmand, and G. L. Doorman. Modelling of prices using the volume in the Norwegian regulating power market. In *2009 IEEE Bucharest PowerTech*, pages 1–7, June 2009.

[9] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[10] G. Klæboe, J. Braathen, A. L. Eriksrud, and S.-E. Fleten. Day-ahead market bidding taking the balancing power market into account. 30(3):683–703. ISSN 1863-8279.

[11] G. Klæboe, A. L. Eriksrud, and S.-E. Fleten. Benchmarking time series based forecasting models for electricity balancing market prices. *Energy Systems*, 6(1):43–61, Mar. 2015. ISSN 1868-3975.

[12] P.-H. Kuo and C.-J. Huang. An Electricity Price Forecasting Model by Hybrid Structured Deep Neural Networks. *Sustainability*, 10(4), 2018. doi: 10.3390/su10041280.

[13] J. Lago, G. Marcjasz, B. De Schutter, and R. Weron. Forecasting day-ahead electricity prices: A review of state-of-the-art algorithms, best practices and an open-access benchmark. 293:116983. ISSN 0306-2619.

[14] A. Mousavian, D. Anguelov, J. Flynn, and J. Kosecka. 3D bounding box estimation using deep learning and

geometry. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pages 7074–7082, 2017.

[15] Optimeering AS. AI apps for power markets. https://optimeering.com/fab-apps, 2022.

[16] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, and *et al.* Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[17] T. S. Salem, K. Kathuria, H. Ramampiaro, and H. Langseth. Forecasting Intra-Hour Imbalances in Electric Power Systems. *Proceedings of the AAAI Conference on Artificial Intelligence*, 33(01):9595–9600, Jul. 2019.

[18] K. Skytte. The regulating power market on the Nordic power exchange Nord Pool: an econometric analysis. *Energy Economics*, 21(4):295–308, Aug. 1999. ISSN 0140-9883.