Short Term Net Imbalance Volume Forecasting through Machine and Deep Learning: A UK case study [⋆]

Elpiniki Makri, Ioannis Koskinas, Apostolos C. Tsolakis[0000−0003−2606−1402], Dimosthnenis Ioannidis[0000−0002−5747−2186], and Dimitrios Tzovaras[0000−0001−6915−6722]

Information Technologies Institute, Centre for Research and Technology Hellas, 6th km Harilaou-Thermis, GR57001, Thessaloniki, Greece
{elpimak, jkosk, tsolakis, djoannid, Dimitrios.Tzovaras }@iti.gr
https://www.iti.gr

Abstract. As energy markets become more and more dynamic, the importance of price forecasting has gained a lot of attention over the last few years. Considering also the introduction of new business models and roles, such as Aggregators and energy flexibility traders, in the constantly evolving energy landscape which follows the general opening of the European electricity markets, the need for anticipating energy price trends and flows holds significant business value. On top of that, the exponential renewable energy sources penetration, adds to the challenges introduced to this dynamic scheme of things. Given their volatile and intermittent nature, supply-demand imbalance can reach critical margins, threatening the overall system stability. In the scope of reducing the power imbalances, a forecast for the imbalance volume will be beneficial either from the perspective of the system operator that could minimise mitigation costs, or the market participants that could target extreme prices for maximising their profit, while effectively managing their risks. The development of a deep learning algorithm for the prediction of the net imbalance volume in the UK market is proposed in this paper in comparison with a common but widely used machine learning approach, namely a gradient boosting trees regression model. The variables which contributed the most on those models were mainly the historical values of net imbalance volume. The deep neural network returns a Root mean squared error (RMSE) and Mean Absolute Error (MAE) equal to 200 and 152 MWh in a range of values between [-1.5,2.0] GWh, respectively, the gradient boosting trees model has an RMSE and MAE equal to 203 and 154 MWh, in contrast to an ARIMA model that has RMSE and MAE equal to 226 and 173 MWh.

Keywords: Net Imbalance Volume · Time Series Forecasting · Long Short-Term Memory · Gradient Boosting trees · Machine Learning · Energy Forecasting · Imbalance Forecasting.

1   Introduction

In recent years the attention of energy's community has turned to Renewable Energy Sources(RES) in the direction of clean energy generation and less carbon emissions. However, the incorporation and provision of intermittent energy to the electric grid tends to create grid stability issues or high volatility in the energy prices. Thus, the confrontation of that kind of issues from National Grid Energy System Operator (NGESO) or aggregator's side, demand efficient forecasting tools for mitigation of unexpected net imbalance deviations.

In 2020, information regarding the electricity sector's grid supply for the United Kingdom (UK) came from 55% low-carbon power (including 24.8 % from wind, 17.2 % nuclear power, 4.4% solar, 1.6% hydroelectricity, 6.5% biomass), 36.1% fossil fuelled power and 8.4% imports . Fuel-based generation and in particular coal-based generation, in contrast to its former domination, nowadays is mainly employed during winter due to pollution and high operational cost [5]. While coal generators have a downward trajectory, renewable power takes the lead and seems to keep constantly growing. By February 2018, UK held the world's sixth place of the largest producers of wind power, with 12,083 megawatts of onshore capacity and 6,361 megawatts of offshore capacity, which leads to a total installed capacity of over 18.4 gigawatts. Moreover, solar power consists another RES growing fast in UK(the third-largest solar energy producer in Europe in 2018)[2], providing significant generation during the day, but is considered inadequate in terms of total energy provided.

The role of the Transmission System Operator (TSO) is entrusted with the management and development of the grid transmission as also for maintaining a constant balance between electricity supply from power stations and demand from consumers.[8]. Along with TSO, market participants submit their supply for both up-and down-regulation individually. Specifically, the producers urged to be aware of their bids that got approved in the day-ahead market as well the spot price (the amount of electricity the market needs at any moment). During the bids, the participants determine the amount of power and the according value of price it will offered for regulation in each hour of the following day. The market actors that have caused an imbalance in the power market will charged for the balancing reserves that have been activated, in order to restore balance in the power system [9]. Net Imbalance Volume (NIV), which is examined in this paper, represents the volume of balancing actions remaining after the volume of the Buy balancing actions ("Offers"), are netted off against the volume of Sell balancing actions ("Bids"). The NIV as a metric indicates mostly the market participants' response to the modifications in the balancing arrangements, rather than the direct effect caused by the application of those modifications [24].

As dynamic energy markets are relatively new, there are very few findings in the literature regarding imbalance market forecasting, and even more specifically targeting NIV. Imbalance market's Price Forecasting challenge has been addressed from [13][27][20] through probabilistic forecasting models, while [7] denoted the importance of NIV as a highly correlated feature to the price. More specifically in [7], a statistical approach has been followed for the calculation

of a transition state probability of NIV over historical data. Traditional time-series algorithms' such as ARIMA were tested in [10] over univariate data, while also conducted autocorrelation and partial autocorrelation analysis in order to explore association between the time periods, to conclude that Feed Forward Networks over multivariate data can achieve higher accuracy because of the problem's complexity. In [3], an encoder-decoder model was designed for a short term probabilistic forecast that was combined with an optimization algorithm for optimal market participation. From the perspective of density forecasting models, [4] endeavoured to predict the imbalances in Austrian energy markets through the exploitation of historical imbalances, the historical load forecast errors, the wind and the solar production. Finally, as the RES can be considered a significant factor for increased or decreased energy reserves during the day, [11], [15] explored the impact of Wind and Solar generation towards the markets' behaviour and price, identifying high correlations between the weather with NIV volatility ratios and prices.

The presented work aims to address the challenge of accurately forecasting NIV in the UK market, by using AI-driven techniques and combining knowledge acquired from historical information within related energy markets, in order to reduce the imbalances of the market as much as possible and introduce an added value service to market stakeholders for optimally participating in dynamic markets and potentially increasing the revenue streams. The remainder of the paper is structured as follows: Section 2 introduces the methodology and architecture of the deep learning model designed. Section 3 describes the details of the experimental data set, while Section 4 presents the experimental setup and results for the evaluation and validation of the implemented architecture. Finally, Section 5 concludes the manuscript with some key points and suggestions for further work in order to expand the research in this area.

2   Methodology

Time series analysis is a statistical technique that deals with time series data, and in particular is considered as the use of a model to predict future values based on previously observed values. In order to determine a forecasting model a representation and further examination of the historical patterns should be beneficial for the analysis. While analysing historical data through time for identifying any trends/patterns, an assumption is made that the existing trend would continue in the future. During the before mentioned exploration the seasonal variation, other cycle variation (i.e. The four stages of the economic cycle, expansion, peak, contraction, and trough) and irregular fluctuations of the series values could be distinguished. End wise, stationarity denotes that the statistical properties of a process generating a time series is stable over time, inevitably its mean, variance and covariance don't have significant changes over time [21] .

Forecasts contain ambiguities that will irrevocably lead to forecast errors. In the scope of calculating the extend of forecasting error, Mean Absolute Error (MAE) and Mean Squared Error (MSE) are the metrics serving this objective

and are used to evaluate the performance of the model in regression analysis. The Mean absolute error represents the average of the absolute difference between the actual and predicted values. Mean Squared Error represents the average of the squared difference between the original and predicted values. Additionally, another deterministic metric is the square root of Mean Squared error namely the Root Mean Squared Error (RMSE).

Among the numerous methods aimed at achieving accuracy and minimizing losses within time series forecasting, there are several machine learning algorithms that induce high precision and computational efficiency [16]. One important characteristic of NIV forecasting is the structure of the time series of both the input variables and the forecasted output. This transforms the problem of forecasting NIV into a specialized form of regression, thus the predicted outcome matches a numerical or continuous value. Nevertheless, traditional machine learning algorithms are affected relentlessly by the missing values and are governed by the deficiency of recognizing complex pattern. Recurrent Neural Network's (RNN) performance is not significantly affected from missing values, as also RNNs can find complex patterns in the input time series. In addition to standard RNNs, Long Short-Term Memory Networks (LSTM), introduced by Hochreiter Schmidhuber [23], have been developed to overcome the vanishing gradient problem by improving the gradient flow within the network. This is achieved using a LSTM unit in place of the hidden layer [22].

The cell state into the LSTM cells acts as a highway in order for the gradient to flow better to the earlier states, which in turn allows the model to capture memory that are further back in the past. Information is removed or added to the cell state, carefully regulated by structures called gates. Gates are mainly a pointwise multiplication operation and a sigmoid neural net layer, thus the output values are fluctuated between 0 (all of the information is removed) and 1(all of the information passes through).Three gates constitutes the LSTM, the Forget Gate, the Input Gate and the Output Gate. Forget Gate adjusts the amount of previous data information which will pass through, while the Input Gate decides which values will be updated. Output Gate produces the output that is multiplied by a tanh, in order for the output to be filtered [26]. In this paper we compared the performance of a recurrent neural network (Figure 1) and a gradient boosting trees (i.e. using the XGBoost implementation [6]) training algorithm for forecasting the NIV value.

The Figure 1 depicts the architecture of the deep learning neural network which includes 6 LSTM layers. The LSTM model in Keras from Tensorflow [1] framework is defined as a sequence of layers. The first layer in the network defines the 2-dimensional units of the tensors and contains 128 number of neurons. The rest of the layers are stacked by adding them to the sequential process and most specifically the second layer contains 64 units, the third and fourth layer have 128 and 64 respectively and the last two contain 64 and 16 number of units respectively. The activation function used in each layer is the Rectified Linear Units (ReLU)[25], which transforms the output from each unit as the input for
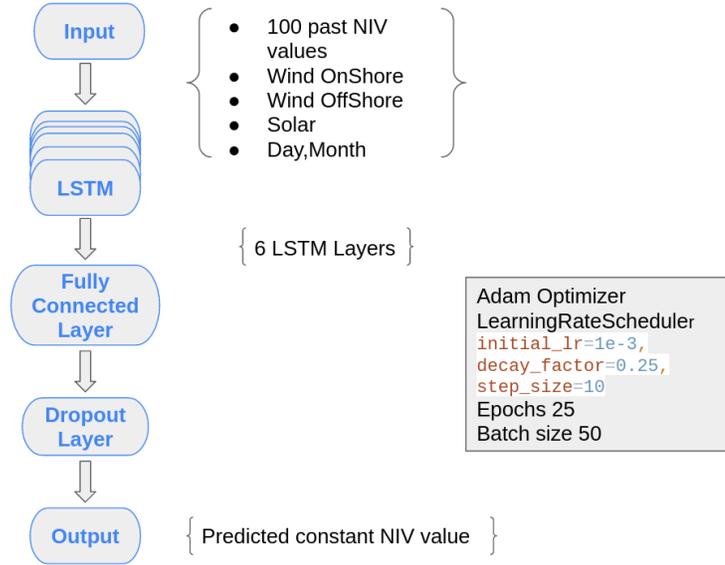
---

[1] https://pypi.org/project/Keras/

Fig. 1. Recurrent Neural Network architecture

the next layer. The selection of ReLU is due to simpler mathematical operations, hence less computational complexity, compared to tanh and sigmoid, as also is widely used in cases where vanishing gradient problems need to be avoided and rectified. Subsequently, a Dense layer is added with number of units equal to the number of features(134). In order to avoid overfitting in such a large network and improve performance, some type of regularization and dropout have subjoined to the model. Regularization reduces parameters and simplifies the model by penalizing high-valued regression coefficients and most specific L2 regularization, which is applied on the first layers of the model, adds an L2 penalty equal to the square of the magnitude of coefficient. Ridge regression (L2 regularization) is frequently used when the independent variables are highly correlated (multicollinearity) and the high value of variance causes the deviation of observed values from the actual values. On the attemp of solving this particular issue, a shrinkage parameter lambda is added [18]. L2 regularization decreases the complexity of a model, however it never leads to a coefficient tending to zero rather only minimizes it. On the other hand, the dropout layer randomly sets input units to zero with a frequency of rate at each step during training time. Several values of rates between 10% to 50% have been applied in order to conclude to the value on which the model's performance is optimized. The 20% value for the dropout layer emerged as the most suitable. Finally, the last layer constitutes of a Dense layer with 1 unit and the activation function on this particular layer is linear. As the details about the topology of the network have been clarified, the following step concerns the optimization algorithm and

the loss function. The MSE value is selected for the loss function and the Adam optimizer was deemed as the most appropriate. In this implementation instead of a fixed value for learning rate, the LearningRateScheduler of Keras is used which reduces the learning rate according to a pre-defined schedule during training. After the above hyper-parameters have been selected after trying different neural networks and observing at each step the loss function performance, the number of epochs(25) and the batch size(50) are identified.

The historical data are dating since January 2015 until June 2020. Except of the NIV values that are processed, further features and the correlation among them are examined. The feature correlation establish the most important pylon in favor of feature selection. Three different techniques have been employed to extract the most suitable features from the NIV data, namely: i) decision forest regression, ii) gradient boosting trees, and iii) permutation importance on top of the prediction model [14]. The Decision Forest regression is a supervised learning method that creates a regression model consisting of an ensemble of randomly trained decision trees. The outputs of each tree in the decision forest is a Gaussian distribution by way of prediction. The algorithm performs an aggregation over the ensemble of trees in order to find a Gaussian distribution closest to the combined distribution for all trees in the model [17]. Gradient Boosting Trees is another method which is used in machine learning in order to create ensemble models. The algorithm constructs each regression tree step by step. Using the predefined loss function it measures the error in each step and correct it in the next one. The prediction model is an ensemble of the weaker prediction models. In the regression problem, boosting creates a series of trees step-wise, and then using an arbitrary differential loss function, selects the optimal tree [17]. Feature permutation importance measures the predictive value of a feature by evaluating the increase of the prediction error which increases in case of feature's unavailability. The algorithm randomly shuffles the features adding noise, in order to avoid the removal of features and the retraining of the regressor [19].

The scope of the paper lays on the exploration of machine and deep learning techniques for the prediction of the NIV values for the next 30 minutes based on the preceding values, over a large dataset of five years. Subsequently, the examination of the various factors, which presumably have a catalytic affect on the balancing power volume, was developed. The investigation took place in order to distinguish the variables that can be used as explanatory variables for the balancing power volume when developing the forecasting model. Since the analysis concerns the behavior of the net imbalance volume through the time field, the primary parameters which are placed under the microscope was the date characteristics. The model predicts the target value for the next half hour, thus factors such as price value or power consumption are unknown and cannot be used as inputs to the model. However, the wind power production forecast, both Onshore and Offshore as well as the Solar power production are known when forecasting the balancing power volume, hence are included in the set of features. The past values of the balancing power volume can also be used as a predictor for the balancing power volume.

3   Experimental Dataset

The Balancing Mechanism Reporting Service (BMRS) API of Elexon [1], which constitutes of programming instructions for participants in order to retrieve BRMS data, is suited for users seeking to access historical data or real-time information. The historic values regarding the target data explored in this work are ranged from January 2015 until June 2020, updated every 30 minutes. Thus, 48 measurements are counted during a day period. In Figure 2, the NIV time series along the above mentioned period is demonstrated.
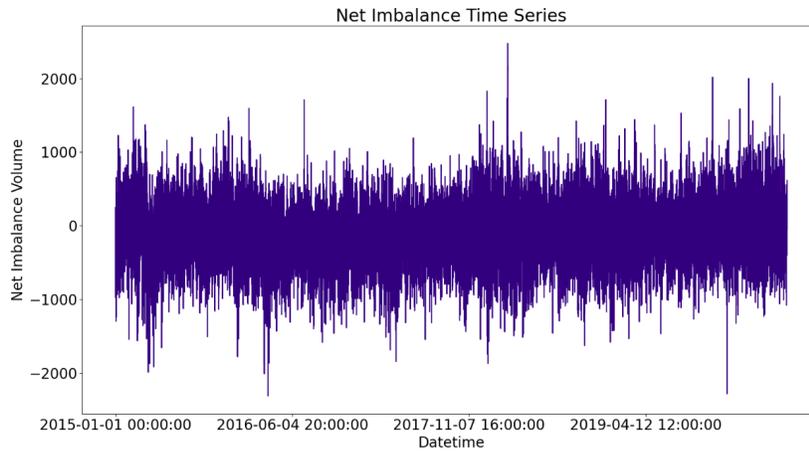

Fig. 2. Net Imbalance Volume Time Series

Evidently, NIV time series are non-stationary whereas the mean as well as the standard deviation are not consistent through time. On the assumption of absent of stationarity, differencing technique overtakes this issue. Nevertheless, this paper conducts with non-stationary time series.

Taking into consideration the increasing amount of power that comes from the renewable sources of energy such as wind and sun, uncertain power consumption have lead to imbalances in the power system. Specifically, based on the weather forecast which entails some uncertainty, wind power producers estimate the amount of power that will produce the following day. In addition, there is a bottleneck regarding the fully estimation of the affect of the heating especially in hours where the sun radiation is high compared to previous hours. Thus, the intensity of the sun radiation is unpredictable. Hence, it is implied that those factors should be examined whether affect the imbalancing volume. Wind, both onshore and offshore, and sun measurements during the day are also available through the (BMRS) API of Elexon. Furthermore, analysing the time

series in order to locate seasonality, the daily and weekly variation of the balancing power will be processed. Firstly, the day has divided to eight groups of four hours, separating the peculiarity of those hours. Then, the date is distinguished to each day of the week and whether it belongs to a working day, as also it is classified to separated months. Thereafter, mean and differencing value are used as candidate features into the feature selection litigation.

The dataset is divided into two sets, the train set and the test set. The dataset is divided by rows, and in order to ensure that the data that has been used for the model training will not be used for the model testing. The first set, which is 70% of the original dataset will be used to train the model, the second set that represents the remaining 30% of the original dataset will be used to generate the predictions. Thus, the data from 2015 to 2019 will be used for the model training and the data from 2019 to 2020 June will be used for the model testing.
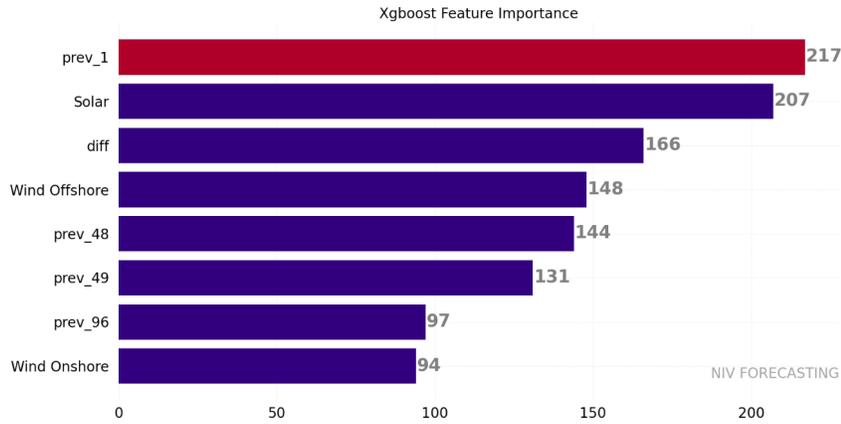
Fig. 3. XgBoost Feature Importance

Conventionally, feature importance algorithms such as feature permutation importance, random Forest decision tree and gradient boosting trees, which extensively described in 2, provide a score which implies the importance of each feature for the model. Figure 3 and Figure 4 depict the outcome of the gradient boosting trees and permutation feature importance algorithms respectively, whereas Figure 5 presents also feature importance calculation results from the the Random Forest technique.

Comparing the images occurred from feature importance analysis, it's clear that both wind and solar power generation affect the forecasted NIV value. Concerning the NIV historical values (100 past values), it seems that one and two days ago values act on the importance estimator, although the previous value
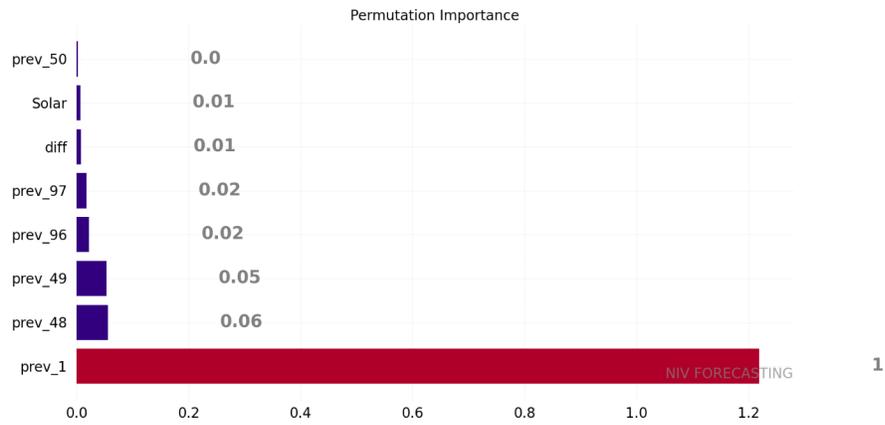
Fig. 4. Permutation Importance
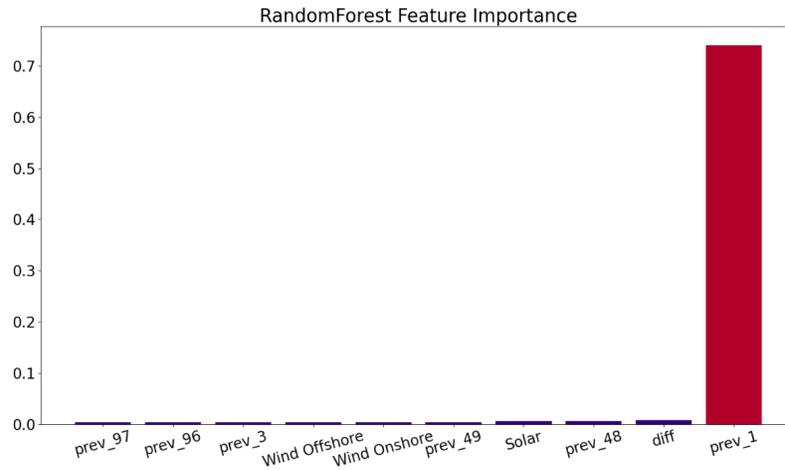
Fig. 5. Random Forest Importance

of the forecasted influences the most. Last but not least, the difference calculated on net imbalance volume plays a part on forecasting performance. Although, the date characteristics are not indicated as the most influencing factors, their use in time series analysis is an indisputable benefit.

4   Results

The Gradient Boosting Tree regression model and the LSTM architecture presented in Section 2 are used in the experiment. Training and testing of each model ran independently and the performance of each model was evaluated. The result from the model evaluation are used in order to compare these models and identify the most suitable one for the given forecasting task and dataset. The accuracy of the models has been evaluated by using the Mean Square Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE) and Symmetric Mean Absolute Percentage Error (sMAPE). In addition with the comparing results, the outcome of an Autoregressive integrated moving average (ARIMA), a widely known statistical analysis model that mainly used in time series data, has been added. The Table 1 shows the various parameters which examined in order to conclude on those that produce better results on forecasting. Regarding the

XGBoost Parameter Tuning	MSE	RMSE	SMAPE	MAE
max_depth = 10 n_estimators = 50	46429	215	96	164
max_depth = 5 n_estimators = 50	41394	203	92	154
max_depth = 5 n_estimators = 100	41605	203	93	154
max_depth = 5 n_estimators =200	42241	205	93	155

Table 1. XGBoost Parameter Tuning

Models	MSE	RMSE	MAE	SMAPE
LSTM	40292	200	152	92
XGBoost	41394	203	154	92
ARIMA	51240	226	173	91

Table 2. Comparative Error Metrics

| Range (MWh)	MSE	RMSE	MAE
0 - 300	27079	164	127
300 - 500	58616	242	193
500 - 2000	147818	384	313

Table 3. Error Metrics for Deep Learning Model in ranges

capability of the models on forecasting the NIV value for next 30 minutes, it can be seen from Table 2 that both of the models produce similar results, although the LSTM is slightly better compared to Gradient Boosting Trees on the same dataset. However, it is clear that both models outperform a simplified approach, such as an Autoregressive integrated moving average (ARIMA). The divergence of the actual value and the forecasted value is within acceptable limits and the models predict the right direction of the imbalance. In fact, compared with previous findings with the same metrics [10], the errors are significantly reduced. Although, it is observed that the models are not able on synchronizing with the peaks in the balancing power volumes for both the up-and down-regulation because of the great influence the previous value has to the model, as already shown in figures 3,4,5. As a result, the models fail to indicate hours where the balancing power volume under the up-or down-regulation will have its peaks. In hours with no regulation the models, in many cases, predict higher values for the up-or down-regulation. This problem might be resolved after applying difference techniques to eliminate the non-stationarity of the NIV time series.

5   Conclusion & Future Work

The aim of this paper was to examine the set of variables which affect the NIV forecasting and design, evaluate and validate the performance of a Deep learning algorithm for predicting short-term future values. NIV describes the necessary energy required to be sold or purchased from the System Operator (SO) towards maintaining the energy system in balance. An accurate forecast can significantly reduce balancing costs, as well as define the necessary preventive mechanisms and early corrective actions to avoid high NIV values. On the other hand, market participants could more efficiently place their bids in the dynamic markets, like the balancing markets, minimizing their risk while maximising their revenues.

From the side of development this paper proposes the implementation of deep learning algorithms which results in slightly better performance than a machine learning counterpart, with the prospect of updating the networks regularly, since the deeper the recurrent network the more data needed in order to preserve lower values in error metrics. The challenges that emerged during the analysis are divided in two main categories. The first one was the missing values in the dataset
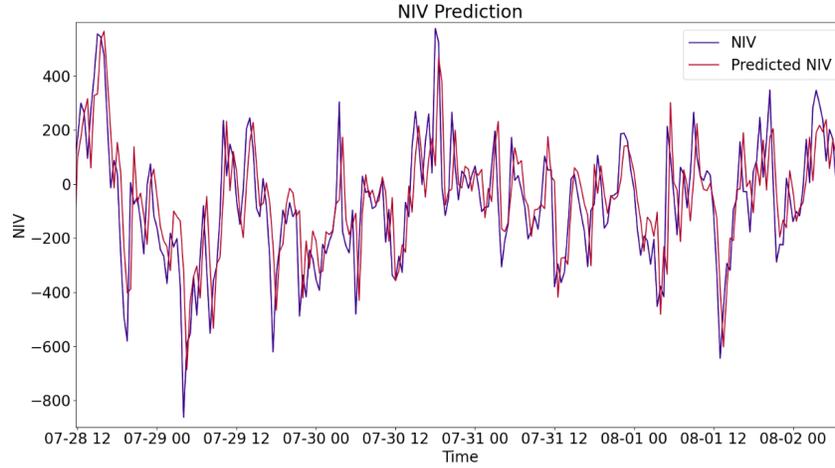
Fig. 6. Actual and forecasted NIV values from LSTM model

as well as the general lack of data to create a highly reliable reference model due to young market conditions. The second one is related to the development of a flexible methodology that can surpass the various parameters which directly affect the market (e.g. unusual market condition such as the incident on September 15,2020, which The UK's electricity system price spiked to over £500/MWh in response to low levels of wind generation [12]).

Despite the fact that the results of the NIV forecasting presented for the next 30 minutes are not optimal -RMSE is equals to 200 MWh- the 63% of the NIV values in dataset have as absolute value less than 300 MWh, which in that case the RMSE is 164. This paper can be the basis for further work and analysis on the balancing market area. Nevertheless, the presented work is one of the first that cover such a long period of data, introducing a more reliable data-driven analysis, whereas compared with previous research endeavours it introduces significantly lower error metrics. Furthermore, the redefinition of the forecasting approach could lead to better results. A suggested approach can be the prediction of an upper and lower value of the balancing volume for the next half hour or the change of the time horizon of prediction to be for the next hour or a day ahead. Additionally, the balancing volume is immediately related with the energy price which regards the spot and forward prices in wholesale electricity markets. Further investigation takes place simultaneously with the forecasting balancing volume also for the energy price, as it is the most crucial predictor factor that needs to examined. On this ground it is suggested for future work the use of forecasting NIV value as a feature for the energy price forecasting.

References

1. Elexon. https://www.elexon.co.uk/, accessed: 2021-03-12
2. Besta, S.: Europe's top five solar power producing countries. NS Energy (2019)
3. Bottieau, J., Hubert, L., De Greve, Z., Vallee, F., Toubeau, J.F.: Very-short-term probabilistic forecasting for a risk-aware participation in the single price imbalance settlement. IEEE Transactions on Power Systems **35**, 1218 – 1230 (03 2020). https://doi.org/10.1109/TPWRS.2019.2940756
4. Bunn, D.W., Gianfreda, A., Kermer, S.: A trading-based evaluation of density forecasts in a real-time electricity market. Energies **11**(10) (2018). https://doi.org/10.3390/en11102658, https://www.mdpi.com/1996-1073/11/10/2658
5. Department for Business, E..I.S.: Uk energy in brief 2020 pp. 31–35 (2020), https://www.gov.uk/government/statistics/uk-energy-in-brief-2020
6. Chen, T., Guestrin, C.: Xgboost: A scalable tree boosting system. In: Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining. pp. 785–794 (2016)
7. Dumas, J., Boukas, I., de Villena, M.M., Mathieu, S., Cornélusse, B.: Probabilistic forecasting of imbalance prices in the belgian context. In: 2019 16th International Conference on the European Energy Market (EEM). pp. 1–7 (2019). https://doi.org/10.1109/EEM.2019.8916375
8. Eduardo A., Martínez Ceseña, P.M.: Smart distribution networks, demand side response, and community energy systems: Field trial experiences and smart grid modeling advances in the united kingdom pp. 275–311 (2018). https://doi.org/https://doi.org/10.1016/B978-0-12-803128-5.00008-8
9. Ferreira, P.P.: Volume and price in the nordicbalancing power market pp. 18 – 21,29–34 (2016), https://core.ac.uk/download/pdf/154676684.pdf
10. Garcia, M.P., Kirschen, D.S.: Forecasting system imbalance volumes in competitive electricity markets. IEEE Transactions on Power Systems **21**(1), 240–248 (2006)
11. Green, R., Vasilakos, N.: Market behaviour with large amounts of intermittent generation. Energy Policy **38**, 3211–3220 (07 2010). https://doi.org/10.1016/j.enpol.2009.07.038
12. Grundy, A.: Uk electricity system price spikes to over £500/mwh due to low wind. Current (2020)
13. Hong, T., Pinson, P., Fan, S.: Global energy forecasting competition 2012. International Journal of Forecasting **30**(2), 357–363 (2014). https://doi.org/https://doi.org/10.1016/j.ijforecast.2013.07.001, https://www.sciencedirect.com/science/article/pii/S0169207013000745
14. Huang, N., Lu, G., Xu, D.: A permutation importance-based feature selection method for short-term electricity load forecasting using random forest. Energies **9**(10), 767 (2016)
15. Ketterer, J.: The impact of wind power generation on the electricity price in germany. Energy Economics **44** (07 2014). https://doi.org/10.1016/j.eneco.2014.04.003
16. Lago, J., De Ridder, F., De Schutter, B.: Forecasting spot electricity prices: Deep learning approaches and empirical comparison of traditional algorithms. Applied Energy **221**, 386–405 (2018)
17. Nagpal, A.: Decision tree ensembles- bagging and boosting. Towards Data Science (2017), https://towardsdatascience.com/decision-tree-ensembles-bagging-and-boosting-266a8ba60fd9

18. Nagpal, A.: L1 and l2 regularization methods. Towards Data Science (2017)
19. Nantian Huang, Guobo Lu, D.X.: A permutation importance-based feature selection method for short-term electricity load forecasting using random forest. Energies (2016). https://doi.org/https://doi.org/10.3390/en9100767
20. Nowotarski, J., Weron, R.: Recent advances in electricity price forecasting: A review of probabilistic forecasting. Renewable and Sustainable Energy Reviews **81**, 1548–1568 (2018). https://doi.org/https://doi.org/10.1016/j.rser.2017.05.234, https://www.sciencedirect.com/science/article/pii/S1364032117308808
21. Palachy, S.: Stationarity in time series analysis. Towards Data Science (2019), https://towardsdatascience.com/stationarity-in-time-series-analysis-90c94f27322
22. Pra, M.D.: Time series forecasting with deep learning and attention mechanism. Towards Data Science (2020), https://towardsdatascience.com/time-series-forecasting-with-deep-learning-and-attention-mechanism-2d001fc871fc
23. Sepp Hochreiter, J.S.: Long short-term memory **9** (1997). https://doi.org/https://doi.org/10.1162/neco.1997.9.8.1735
24. Shah, A.G.A.: Analysis of the first phase of the electricity balancing significant code review. The Office of Gas and Electricity Markets (Ofgem) p. 16 (2018)
25. SHARMA, S.: Activation functions in neural networks. Towards Data Science (2017)
26. Steven Elsworth, S.G.: Time series forecasting using lstm networks: A symbolic approach (2020)
27. Zhang, Y., Wang, J., Wang, X.: Review on probabilistic forecasting of wind power generation. Renewable and Sustainable Energy Reviews **32**, 255–270 (2014). https://doi.org/https://doi.org/10.1016/j.rser.2014.01.033, https://www.sciencedirect.com/science/article/pii/S1364032114000446