Energy Economics 137 (2024) 107770

Contents lists available at ScienceDirect

Energy Economics

journal homepage: www.elsevier.com/locate/eneeco

Seasonality in deep learning forecasts of electricity imbalance prices

Sinan Deng [a,1], John Inekwe [b], Vladimir Smirnov [a], Andrew Wait [a,*], Chao Wang [c]

[a] *School of Economics, The University of Sydney, Social Sciences Building A02, NSW, 2006, Australia*
[b] *Macquarie University, NSW, 2109, Australia*
[c] *Business School, The University of Sydney, Building H70, NSW, 2006, Australia*

A R T I C L E I N F O

JEL classification:
C32
C45
Q41
Q47

Keywords:
Forecasting
Electricity
Balance settlement prices
Deep learning
Machine learning

A B S T R A C T

In this paper, we propose a seasonal attention mechanism, the effectiveness of which is evaluated via the Bidirectional Long Short-Term Memory (BiLSTM) model. We compare its performance with alternative deep learning and machine learning models in forecasting the balancing settlement prices in the electricity market of Great Britain. Critically, the Seasonal Attention-Based BiLSTM framework provides a superior forecast of extreme prices with an out-of-sample gain in the predictability of 11%–15% compared with models in the literature. Our forecasting techniques could aid both market participants, to better manage their risk and assign their assets, and policy makers, to operate the system at lower cost.

1. Introduction

The electricity market is peculiar in that production and consumption need to be in balance in real time, otherwise the entire network can become unstable. In Great Britain, the foundation for the electricity market is bilateral contracts between providers and large consumers. In this arrangement, the role of the system operator (SO) is as a 'balancer of last resort' when there are deviations by market participants from these bilateral obligations.[2] As detailed below, to balance demand and supply in the market in real-time, the SO elicits and accepts the required bids or offers to decrease consumption/increase production (or increase consumption/decrease production) as required, charging or paying an imbalance price to participants. It is these imbalance prices that we forecast in this paper. Specifically, we propose a new deep learning model that includes a seasonal attention mechanism and compare its performance against various parametric techniques and the latest machine learning and deep learning models.

Predicting imbalance prices is becoming increasingly important in the Great Britain electricity market, as penetration of renewables and distributed resources has heightened the reliance on the balancing market by participants (see Borne et al. (2018), Brouwer et al. (2014), Bunn and Kermer (2021) and Hirth and Ziegenhagen (2015)). Indeed, while balancing trades made up just 5 percent of the electricity market in 2012, in 2020 these balancing services regularly exceed 50% of the national demand. An accurate forecast of the imbalance price can help participants manage their assets and mitigate their risk.[3] In other energy prediction fields hybrid BiLSTM models have demonstrated superior accuracy compared with other machine learning or deep learning approaches.[4] Better forecasting is also important for the system as a whole. It can assist policy makers adapt the system as required, given changes in the technology mix. More accurate forecasting can lower the cost of delivering electricity and produce better environmental outcomes by improving investments in infrastructure and generation.

In this study we use the balancing market information from 48 half-hour delivery periods in the British electricity market during 2016–2019, the same data utilized by Bunn et al. (2020) and Lima et al. (2022) in their respective studies of imbalance electricity prices.

* Corresponding author.
E-mail addresses: sinan.deng@sydney.edu.au (S. Deng), john.inekwe@mq.edu.au (J. Inekwe), vladimir.smirnov@sydney.edu.au (V. Smirnov), andrew.wait@sydney.edu.au (A. Wait), chao.wang@sydney.edu.au (C. Wang).

1 The code is available at www.sinandeng.com.

2 In an integrated electricity system, a set of technical and economic institutions used to ensure short-term power balance defines the balancing system, see Hirth and Ziegenhagen (2015).

3 Spot price spikes can contribute significantly to energy prices. Calculating spot price spikes can help with risk management (Lu and Suthaharan, 2023).

4 Noteworthy examples include CEEMD-RF-IRSA-BiLSTM, TVFEMD-RF-CNN-ISCA-BiLSTM and RF-IHHO-CNN-BiLSTM models, as detailed in Xiong et al. (2023), Zhang et al. (2022) and Li et al. (2023).

https://doi.org/10.1016/j.eneco.2024.107770
Received 10 August 2023; Received in revised form 1 July 2024; Accepted 6 July 2024
Available online 14 July 2024

Table 1
List of abbreviations.

ANN: Artificial Neural Network	MLP: Multilayer Perceptron
BiLSTM: Bidirectional Long Short-Term Memory	RF: Random Forest
CEEMD: Complementary Ensemble Empirical Mode Decomposition	IHHO: Improved Harris hawk Optimization
CNN: Convolutional Neural Network	SA-BiLSTM: Seasonal Attention Bidirectional Long Short-Term Memory
GRU: Gated Recurrent Units	SVR: Support Vector Regression
IRSA: Improved Reptile Search Algorithm	TCN: Temporal Convolutional Networks
ISCA: Improved Sine and Cosine Algorithm	TVFEMD: Time-varying Filter based Empirical Mode Decomposition
LightGBM: Light Gradient-Boosting Machine	XGBoost: eXtreme Gradient Boosting
LSTM: Long Short-Term Memory	

Our paper makes two main contributions to the literature. First, we propose a Bidirectional Long Short-Term Memory (SA-BiLSTM) model to forecast UK imbalance prices, allowing for a seasonal attention mechanism that explicitly accounts for important components such as seasonality in the input sequence. Second, we evaluate the forecasting performance of imbalance prices using 11 state-of-the-art machine learning and deep learning models.[5]

Empirical results demonstrate the effectiveness of the deep learning models in the given forecasting task with an out-of-sample gain in the predictability of about 11%–15%, compared with linear, nonlinear, and Bayesian models used in the two existing studies of Bunn et al. (2020) and Lima et al. (2022). In particular, the SA-BiLSTM framework has more predictive power in comparison with other models. Also, critical for both suppliers and consumers, the SA-BiLSTM model provides improved forecasts against extreme prices compared with other frameworks. This could enable market participants to manage risk and opportunity, and optimally assign their assets. Extreme prices, after all, are what generate some risks as well as potentially profitable opportunities for market players.

2. Literature on forecasting in balancing markets

Several recent papers forecast imbalance prices in electricity markets in various jurisdictions. The most relevant for this study are Bunn et al. (2020), Lucas et al. (2020), and Lima et al. (2022) who each focus on predicting imbalance prices in the British balancing market.[6] Bunn et al. (2020) employ a nonlinear technique and reveal the regime-switching behaviour of system imbalance price. Their model indicates that fundamental drivers of wind, solar and demand forecast errors, lagged prices, scarcity variables, and lagged imbalance volume are relevant determinants of imbalance prices. They find that imbalance prices are predictable, and that the out-of-sample predictive power of a regime-switching model is superior to a linear benchmark.

Lucas et al. (2020) argue that, in general, electricity markets are thought to have quasi-deterministic principles; hence it is desirable to employ variables that can describe the outcome of the market when undertaking a forecast of the price. They focus on a multi-variable regression model to explain price behaviour. Their analysis uses three machine learning algorithms: Gradient Boosting (GB), Random Forest (RF), and XGBoost. They find a higher performance of XGBoost, which returns a mean absolute error (MAE) of 7.89 £/MWh. The most significant contributors in their model are the net imbalance volume, the loss of load probability (LOLP), the month and the de-rated margins.

A recent study by Lima et al. (2022) assesses the time-varying nature of the exogenous drivers of the system imbalance prices. Their most accurate forecasts use Bayesian time-varying predictive density method, where the densities are from the Bayesian dynamic linear model, compared with the less successful AR-GARCH model.

Utilizing the key insights of these papers, such as the penetration of renewables and the time-varying impact of exogenous factors, we advance the literature by introducing a new deep learning model to provide a forecast of price variations in the balancing market. Following Bunn et al. (2020), Lucas et al. (2020), and Lima et al. (2022) we focus on imbalance prices in the electricity market in Great Britain. We also draw on the insights of other studies; Bueno-Lorenzo et al. (2013) suggest that the impact of wind energy is an important determinant of imbalance prices; Garcia and Kirschen (2006) emphasize the role of market variables; and Lisi and Edoli (2018) include lagged imbalances and loads in their forecasts. Ocker and Ehrhart (2017) argue that bids of German generator are orientated towards previous auction prices and that they coordinate on a price level which is (far) above the competitive level.

A related literature focuses on forecasting imbalance volumes.[7] Garcia and Kirschen (2006) compare the performance of AutoRegressive Integrated Moving Average (ARIMA), exponential smoothing and neural networks for predicting volume imbalances in the British market. They find that while neural networks have reasonable accuracy in predicting net imbalance volume (NIV) for both short-and long-term horizons, a new network is required for each forecasting situation. Also for Great Britain, Taylor (2006) compares NIV density forecasting and finds that the seasonal ARMA model or a periodic AR model performs best in point forecasting, while the volatility forecasting method shows the contradictory ranking of methods. They conclude that seasonal ARMA point forecasts with forecast error variance generated from a simplistic approach incorporating intra-day seasonality in the variance produce the best result for constructing density forecast.

Other studies examine forecasting electricity imbalance volumes in different jurisdictions. Using least squares regression, quantile regression and autoregressive moving averages (Goodarzi et al., 2019) find that an increase in the absolute values of imbalance volumes in the German electricity market is due to higher wind and solar forecast errors; the spillover of this effect can be witnessed in the form of higher spot electricity prices. Bunn et al. (2018) provide a forecast of Austrian imbalance volumes by developing a full-density model using a skewed-t specification. Bueno-Lorenzo et al. (2013) examine the impact on imbalance volumes from wind energy in Spain and Lisi and Edoli (2018) focus on zonal imbalance markets in Italy.

3. Data and the balancing system price

The British Balancing Mechanism operates on 30-minute intervals ‘Settlement Periods’ (SP) during which the delivery of forward commitments occurs, and consequently, imbalance volumes are settled. All participants must notify the SO of their expected physical positions arising from their forward bilateral contracts an hour before each SP; this point in time is known as the ‘gate-closure’. As part of its role to ensure system stability, the SO must be aware of the expected

[5] These models are: transformer, SVR, XGBoost, LightGBM, MLP, GRU, ANN, LSTM, BiLSTM, SA-BiLSTM and TCN, see Table 1.

[6] Also see, for example, Klæboe et al. (2015) for forecasts of balancing prices in the Nordic system, Mureddu and Meyer-Ortmanns (2018) for Italy and Bunn and Kermer (2021) for an examination of statistical arbitrage in the Austrian market.

[7] For analysis of the design of balancing markets see Kristiansen (2007), Marneris et al. (2019), Möller et al. (2011), Muratori and Rizzoni (2015), Poplavskaya and de Vries (2019), Vandezande et al. (2010), Szabó et al. (2010) and Wu et al. (2020).

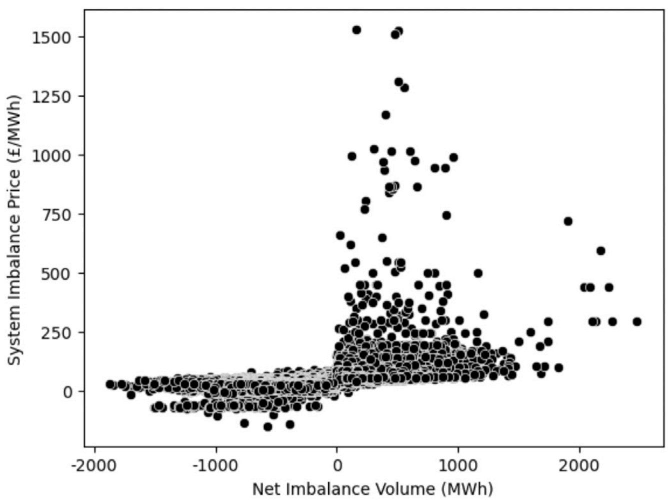

Fig. 1. System Imbalance Price (£/MWh) against NIV.

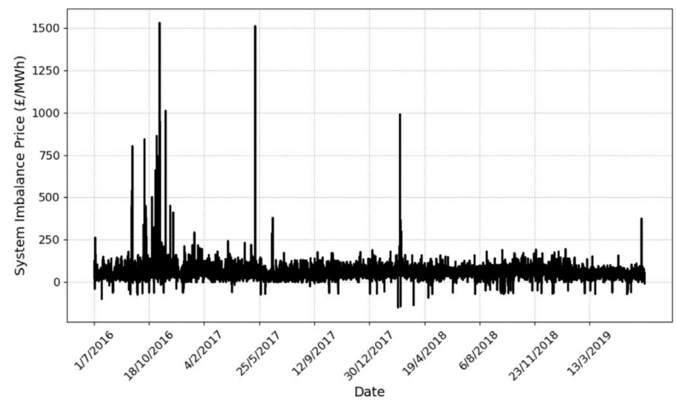

Fig. 2. System Imbalance Price (£/MWh).

amount of electricity to be generated or consumed by all participants during that half-hour. Prior to gate-closure, all flexible generators (or demand-side suppliers/aggregators) can voluntarily notify the SO of their offers/bids to increase/reduce generation (or reduce/increase demand). Along with the price/quantity offers or bids, market participants must also inform the SO of their physical capabilities of altering their generation or consumption.

Typically, there will be deviation by some market participants from their stated physical notification, for a variety of reasons. The SO uses the Balancing Mechanism, which is akin to an auction, as one tool to ensure the system remains in balance in real time; the SO attempts to balance the market in the most cost-effective way possible. In practice, controlling the balancing requires the acceptance of many offers and bids on a minute-by-minute basis (participants must follow these directives).

In any SP, if the transmission system is *long* – that is, there is too much electricity – the imbalance price is based on actions taken by the SO to reduce generation or increase demand. Conversely, when the transmission system is *short* there is insufficient electricity given the quantity demanded. The imbalance price in a short market is calculated on the basis of the SO's actions to increase the generation of electricity and/or decrease the quantity demanded.[8]

Accounting for their participation in the Balancing Mechanism, if a participant has under-generated (or over-consumed for suppliers) in an SP relative to its contracted volume, it will have to buy the shortfall at the imbalanced price from the SO. If, on the other hand, a participant has over-generated (or under-consumed for suppliers) electricity compared with its contracted volume, it will have to sell the extra amount at the imbalanced price.

The data used in this study covers the period between 1st July 2016 and 30th June 2019. It includes information on 48 half-hour delivery periods for each day in the period for the British electricity market. The origin of the data is the Balancing Mechanism Reporting Service (BMRS) provided by ELEXON. To aid comparisons with the existing literature, we use the data previously analysed by Bunn et al. (2020) and Lima et al. (2022). As the SP comprises 48 data points per day, there are 52 560 data points over the three years.

The variables are De-rated Margin, Demand Error, Solar Error, Wind Error, Inter Delta, Non-Balancing Mechanism-NonBM, Net Imbalance Volume-NIV, and System Price. De-rated Margin (MW) reflects the discrepancy between the combined generation forecast and the capacity requirement for the market. It is an inverse measure of scarcity. The error in the day-ahead national demand forecast is the Demand Error, while the error in the day-ahead solar generation forecast is the Solar Error (MW). Wind Error (MW) reflects the error in the day-ahead wind generation forecast. Inter delta reflects the short-term change in power inflows from interconnections outside the UK. The Non-Balancing Mechanism measures the volumes outside of the balancing mechanism. It is the extra reserve power sourced by the SO within each SP.

Net Imbalance Volume (MWh) is the overall system energy imbalance at each SP through the summation of all energy balancing buy and sell actions. If a negative value is recorded, a power surplus should occur in the market and vice versa. System price (£/MWh) is the indicator for imbalance price. As noted in Bunn et al. (2020) and also applied in Lima et al. (2022), occasional negative prices and some spikes occur in the series. Therefore, the series are trimmed slightly (replacement of values above 140 with 140 and negative prices with zeros). While there might be pros and cons for such trimming, the same data pre-processing is followed here to directly compare results with the two existing studies. Importantly, the time series statistical properties of the imbalance system price are sustained while modelling the behaviour of the series. Moreover, the predictability of prices is not marred by such outliers.

The representation of the association between system price and net imbalance is plotted in Fig. 1. System imbalance represents the marginal cost to the SO in each SP. The imbalance volume captures the difference between ex-ante nomination and ex-post-metred volumes in the SP. Thus, if the system is short, the net imbalance volume (NIV) will be positive as the SO will accept more offers than bids and vice versa. The system imbalance price will be the marginal bid paid at a negative NIV or the marginal offer accepted at a positive NIV. The imbalance price will be higher than the marginal cost if a system is short (NIV > 0) and vice versa. In essence, the marginal generation cost of participants will be lower than offers, and the marginal generation cost will be higher than bids; otherwise, generators will lack the interest in increasing or reducing output by placing those offers and bids (see, Bunn et al. (2020)).

Fig. 2 presents the plot of the system imbalance price. The final data is slightly trimmed following the existing studies.

4. Methodology

This section first reviews the LSTM and BiLSTM models. We then introduce the seasonal attention mechanism, which provides extra focus on seasonalities and cycles in the input sequence. The proposed seasonal mechanism is employed in the BiLSTM (to form the SA-BiLSTM framework).

[8] There are two energy prices in each SP, the System Buy Price (SBP) and the System Sell Price (SSP), however since 2015 the SBP will equal the SSP in any SP. See https://bscdocs.elexon.co.uk/guidance-notes/imbalance-pricing-guidance for more details on the Balancing Mechanism.

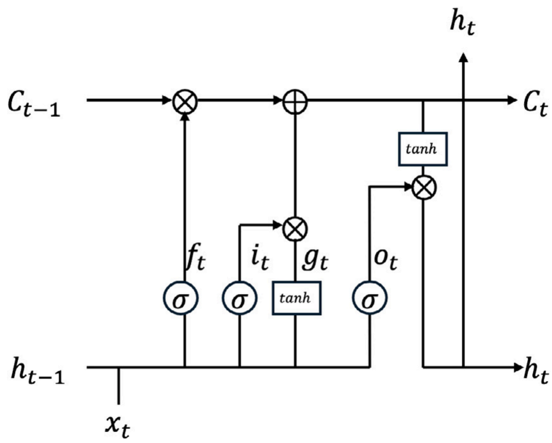

Fig. 3. The LSTM Model.

4.1. The LSTM model

In time-series forecasting, there are two main approaches to capture time effects. One is to represent time effects explicitly via some functions of lagged values, such as the ARIMA model. The other is to capture time effects implicitly via latent variables to store memories of the data dynamics. These latent variables, which are also called hidden states, are updated in a recurrent manner using the information from previous and current time steps. The LSTM neural network evolved from the RNN neural network (Hochreiter and Schmidhuber, 1997). It improves the basic RNN by introducing three extra hidden units: the input gate, the forget gate and the output gate. Fig. 3 shows the LSTM model structure and the LSTM cell can be represented as

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad \text{Forget Gate} \tag{1}$$
$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad \text{Input Gate} \tag{2}$$
$$g_t = tanh(W_g \cdot [h_{t-1}, x_t] + b_g) \quad \text{Candidate Cell State} \tag{3}$$
$$C_t = f_t \odot C_{t-1} + i_t \odot g_t \quad \text{Cell State Update} \tag{4}$$
$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad \text{Output Gate} \tag{5}$$
$$h_t = o_t \odot tanh(C_t) \quad \text{Hidden State} \tag{6}$$

where: f_t, i_t, g_t, o_t are respectively the output values of the forget gate, input gate, candidate cell state and output gate; W_f, W_i, W_g, W_o are respectively the weights for forget gate, input gate, candidate cell state and output gate; b_f, b_i, b_g, b_o are the biases for forget gate, input gate, candidate cell state and output gate, respectively; C_t is the cell state update and C_{t-1} is the previous cell state; x_t is the current input; $tanh$ is the hyperbolic tangent activation function; σ is the sigmoid activation function; $\odot$ is element-wise multiplication; h_t is the hidden state; and h_{t-1} is the previous hidden state.

4.2. The BiLSTM model

A Bidirectional LSTM (BiLSTM) allows the input flows in both directions and utilizes the information from both sides. In other words, the BiLSTM Model is a two-layer LSTM Model with a forward LSTM layer and a backward LSTM layer. The backward LSTM layer is designed in a similar manner to the normal forward LSTM layer, except the direction is reversed to capture subsequent time information. The structure of the BiLSTM model is shown in Fig. 4 and it can be expressed as

$$\overrightarrow{h}_t = \mathrm{LSTM}_{\mathrm{forward}}(x_t, \overrightarrow{h}_{t-1}) \quad \text{Forward LSTM} \tag{7}$$
$$\overleftarrow{h}_t = \mathrm{LSTM}_{\mathrm{backward}}(x_t, \overleftarrow{h}_{t+1}) \quad \text{Backward LSTM} \tag{8}$$
$$y_t = [\overrightarrow{h}_t, \overleftarrow{h}_t] \quad \text{Combined Hidden State} \tag{9}$$

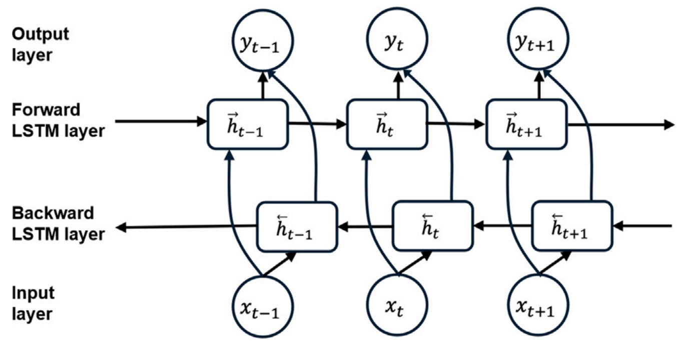

Fig. 4. The BiLSTM Model.

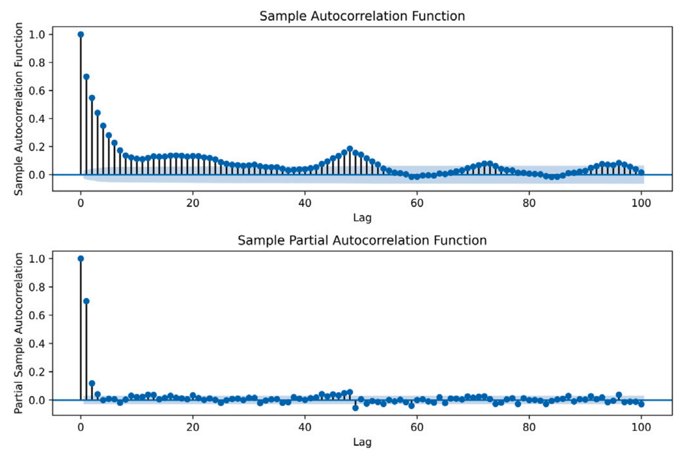

Fig. 5. Sample ACF and PACF for System Imbalance Price.

$$\hat{y}_{t+k} = W_y y_t + b_y \quad \text{Fully Connected Layer} \tag{10}$$

where: $\overrightarrow{h}_t$ is the output from the forward LSTM; $\overleftarrow{h}_t$ is the output from the backward LSTM; y_t is the concatenated vector comprising the hidden state from the forward LSTM $\overrightarrow{h}_t$ and the hidden state from the backward LSTM $\overleftarrow{h}_t$; x_t is the input at t; W_y is the weight for $\hat{y}_{t+k}$ and b_y is the bias for $\hat{y}_{t+k}$.

4.3. The SA-BiLSTM model

In this section we introduce the Seasonal Attention-Based Bidirectional Long Short-Term Memory (SA-BiLSTM) model. One novelty of the model is that it strengthens the learning of seasonal dependency on a daily basis in the input sequence. The SA-BiLSTM is formulated to forecast imbalance electricity prices. Fig. 5 presents the sample autocorrelation function (ACF) and partial autocorrelation function (PACF) for the system imbalance price. The ACF values decay slowly, and around the first 50 lags are outside of the 95% confidence band. Also, with seasonality present, both ACF and PACF have spikes at the seasonal lags with multiples of 48. This is somewhat intuitive, as the system imbalance price at 11 am today might also imply some useful information for the price at 11 am tomorrow, for example.

Fig. 6 presents the framework of the proposed SA-BiLSTM model in detail. Mathematically, the seasonal attention layer can be written as

$$x = [x_1, x_2, \dots, x_{48}] \tag{11}$$
$$W_a = [\omega_1, \omega_1, \dots, \omega_{48}] \tag{12}$$
$$x_t = x \cdot W_a^T \tag{13}$$

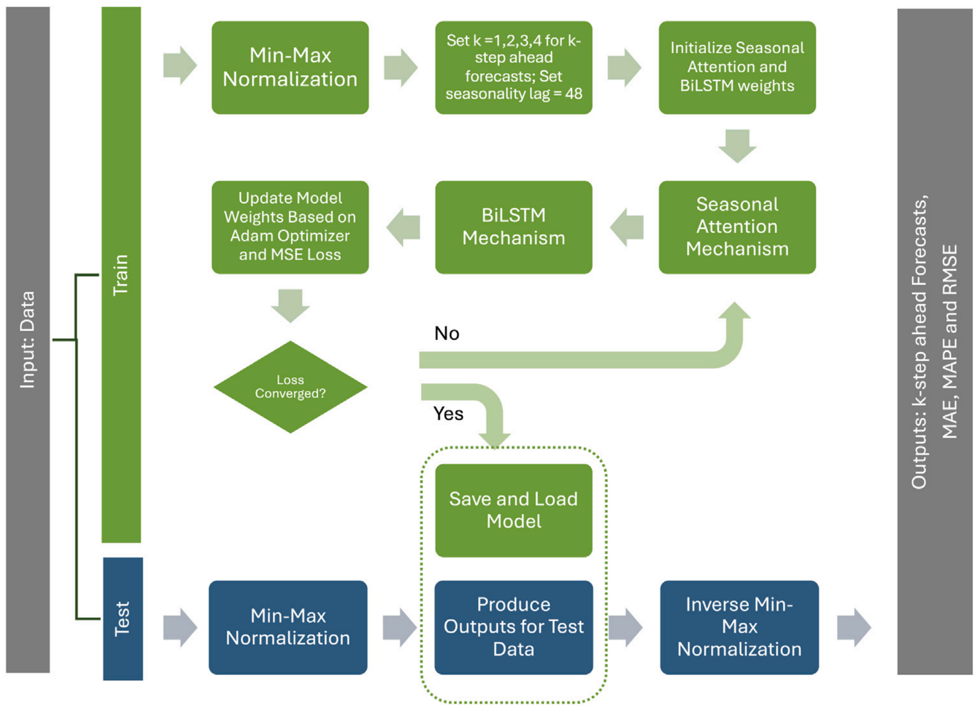

Fig. 6. Flowchart of the Proposed SA-BiLSTM Model.
Note. Grey represents input and output data. Green denotes the training process. Blue denotes testing. The data is separated into the training and test sets (the final 2000 data points utilized for testing). The Min–Max Scaler is used to normalize the training and test data separately, using only values from the training set to prevent data leakage. Initial weights for the attention and the BiLSTM mechanisms are updated during training to ensure loss convergence. The resulting model is applied to the test data, generating output predictions, then reconverted to its original scale. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

where: x is a vector which contains 48 lags of the data; and W_a is a vector of the seasonal attention layer weights.[9]

We also examined 10 other machine learning and deep learning models' ability to forecast imbalance prices. Comparing our model against other models allows for a comparison and can be used as benchmarks (Ghoddusi et al., 2019). These include long short-term memory (LSTM), Bidirectional LSTM (BiLSTM)-(Schuster and Paliwal, 1997), Artificial Neural Network (ANN)- (Mirakyan et al., 2017; Shambora and Rossiter, 2007), Support Vector Regression (SVR)- (Vapnik, 1999; Papadimitriou et al., 2014), Extreme Gradient Boosting (XGBoost)- (Friedman, 2002), Multilayer Perceptron (MLP)- (Thomaidis and Dounias, 2012), Gated Recurrent Unit (GRU)-(Cho et al., 2014), Light Gradient-Boosting Machine (LightGBM)-(Ke et al., 2017), Temporal Convolutional Networks (TCN)- (Bai et al., 2018) and transformer (Vaswani et al., 2017). Machine learning and deep learning models have become helpful in predicting demands and financial asset prices. The utility of these models arises from their ability to make accurate predictions, notably where there are non-linearities in time series data.

5. Empirical analysis

This section aims to evaluate the various models of machine learning and deep learning against the benchmark study (Bunn et al., 2020), as well as the machine learning benchmark SVR Model. First, we measure the accuracy of predictions when compared to parametric models. We also evaluate our proposed model's ability to predict the imbalance of electricity prices over the various windows and across varying periods. The measures of performance include Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Root Mean Squared Error (RMSE). The results are broadly consistent across these indicators, and we focus the interpretation on RMSE. Besides these evaluation methods, we also visualize the mean forecast errors from Table 3 in Figs. 7 and 8. Fig. 7 presents the relative mean squared error decomposition; Fig. 8 shows the relative cumulative mean squared errors.

5.1. Forecasting performance of MAE, MAPE and RMSE

Our entire sample period covers observations from 1st July 2016 to 30th June 2019 with 52,560 data points. However, we do recognize that the data in 2019 is less volatile than the data in 2016. Therefore, in establishing the performance of machine learning and deep learning models and our proposed model, we employ two forecasting windows that are comparable to the application in the benchmark study (Bunn et al., 2020) in this section. Table 2 is obtained by using the entire sample period, while Table 3 is obtained by using a sub-sample in 2019. In each of the forecast windows, one-step to four-step ahead forecasts are produced for each of the 11 machine learning and deep learning methods.

In Table 2, the sample period covers observations from 1st July 2016 to 30th June 2019 with 52,560 data points. The final 2000 data points are used for testing purposes and the rest is used for training. The out-of-sample predictive ability of different models is presented. The out-of-sample performance of these models shows a superior forecast for SA-BiLSTM, with an RMSE of 12.709 for one-step ahead forecast. The performance of the models from the best to the least is given as SA-BiLSTM, LSTM, GRU, BiLSTM, TCN, SVR, MLP, Transformer, ANN, XGBoost and LightGBM. Considering a two-step ahead forecast, the performance of the models follows this order from best to the least: SA-BiLSTM, GRU, LSTM, BiLSTM, SVR, Transformer, TCN, MLP, XGBoost, ANN and LightGBM. The SA-BiLSTM performs better than the tested alternatives with an RMSE of 15.216. For a three-step ahead

[9] Other key parameters of the proposed model are as follows. The number of evolution iterations is fixed at 10 while the batch size is set to 128. The initial epochs are configured as 100, and 10 epochs are set for competitive search. The number of hidden neurons is set at 64 for a double layer architectural design, and the learning rate parameter is set at 0.00001.

Table 2
MAE, MAPE, and RMSE for 2016–2019 Out-of-Sample Forecasts.

(a) Step 1

Method	MAE	MAPE	RMSE
SA-BiLSTM	9.257	37.326	12.709
LSTM	9.215	38.372	12.875
GRU	9.279	40.113	12.894
BiLSTM	9.225	40.128	12.963
TCN	9.356	42.527	13.454
SVR	10.310	47.439	13.476
MLP	9.810	44.546	13.513
Transformer	9.369	42.289	13.553
ANN	10.242	47.712	13.978
XGBoost	10.726	60.164	14.346
LightGBM	15.349	98.883	18.845

(b) Step 2

Method	MAE	MAPE	RMSE
SA-BiLSTM	11.595	51.830	15.216
GRU	11.704	52.766	15.450
LSTM	11.817	52.624	15.465
BiLSTM	11.753	53.995	15.477
SVR	12.429	58.231	15.715
Transformer	11.860	55.930	15.824
TCN	12.154	58.097	16.067
MLP	12.297	41.970	16.428
XGBoost	12.958	73.998	16.834
ANN	13.738	71.096	17.412
LightGBM	15.998	101.461	19.611

(c) Step 3

Method	MAE	MAPE	RMSE
SA-BiLSTM	13.078	61.978	16.516
MLP	13.038	63.296	16.632
BiLSTM	13.296	64.936	16.853
LSTM	13.234	62.440	16.915
GRU	13.249	63.016	16.922
SVR	13.706	65.537	16.993
Transformer	13.389	64.110	17.367
TCN	13.873	68.940	17.597
XGBoost	14.333	83.121	18.265
ANN	15.313	83.321	19.173
LightGBM	16.458	103.478	20.134

(d) Step 4

Method	MAE	MAPE	RMSE
SA-BiLSTM	13.759	68.836	17.169
LSTM	13.831	68.962	17.235
MLP	13.615	62.585	17.300
GRU	14.035	69.794	17.558
SVR	14.351	70.010	17.561
BiLSTM	14.159	72.093	17.589
TCN	14.772	75.404	18.475
Transformer	14.948	81.016	18.783
XGBoost	14.993	88.980	18.853
ANN	16.219	92.464	20.112
LightGBM	16.672	105.242	20.437

Note. The sample period is July 1 2016 to June 30 2019 (52560 data points, with the final 2000 used for testing). k=1,2,3,4 step(s) ahead forecasts are produced for the 11 machine learning and deep learning models.

Table 3
MAE, MAPE, and RMSE for 2019 Out-of-Sample Forecasts.

(a) Step 1

Method	MAE	MAPE	RMSE
SA-BiLSTM	9.321	40.412	13.128
GRU	9.577	42.919	13.181
ANN	9.712	43.564	13.328
XGBoost	10.818	66.324	14.374
LSTM	11.430	56.670	14.815
MLP	11.543	52.154	15.007
SVR	12.625	71.496	15.863
Transformer	12.879	66.801	16.303
BiLSTM	14.272	78.197	16.693
TCN	13.552	61.605	17.589
LightGBM	14.924	89.226	17.944

(b) Step 2

Method	MAE	MAPE	RMSE
SA-BiLSTM	12.098	57.294	15.601
GRU	12.115	59.385	15.776
ANN	12.313	58.898	15.902
LSTM	12.931	65.964	16.560
XGBoost	13.165	77.886	16.733
MLP	13.517	64.473	17.236
SVR	13.892	78.965	17.239
Transformer	14.150	76.262	17.750
BiLSTM	15.512	86.026	18.329
LightGBM	15.451	91.074	18.502
TCN	15.673	85.978	19.187

(c) Step 3

Method	MAE	MAPE	RMSE
SA-BiLSTM	13.489	65.812	16.839
GRU	13.672	69.318	17.203
ANN	13.703	68.494	17.243
MLP	13.631	65.766	17.685
SVR	14.397	81.570	17.746
LSTM	14.329	77.358	17.777
XGBoost	14.401	84.165	17.790
BiLSTM	15.519	85.441	18.546
Transformer	15.145	83.683	18.628
LightGBM	15.846	92.822	18.872
TCN	15.869	87.282	19.264

(d) Step 4

Method	MAE	MAPE	RMSE
SA-BiLSTM	14.180	71.016	17.501
GRU	14.488	77.280	17.981
ANN	14.654	76.669	18.011
SVR	14.765	83.305	18.104
MLP	14.769	80.119	18.525
LSTM	15.072	81.658	18.545
XGBoost	15.111	89.268	18.569
Transformer	15.215	79.279	18.730
TCN	15.234	77.691	18.852
LightGBM	16.117	94.717	19.210
BiLSTM	16.283	92.809	19.295

Note. The sample period is the final 4396 observations in 2019 (the last 2000 data points used for testing). k=1,2,3,4 step(s) ahead forecasts are produced for each of the 11 machine learning and deep learning models.

forecast, the models perform from the best to the least as follows: SA-BiLSTM, MLP, BiLSTM, LSTM, GRU, SVR, Transformer, TCN, XGBoost, ANN and LightGBM. The SA-BiLSTM performs better than the tested alternatives with an RMSE of 16.516. A similar result is obtained for the four-step ahead forecast. The models perform from the best to the least as follows: SA-BiLSTM, LSTM, MLP, GRU, SVR, BiLSTM, TCN, Transformer, XGBoost, ANN and LightGBM. The proposed SA-BiLSTM has an RMSE of 17.169.

In Table 3, the final 4396 observations in 2019 are selected, and the first 2396 observations of these are used in training and estimating the model before forecasting one period ahead throughout the final 2000 out-of-sample observations as done in the benchmark study. The out-of-sample predictive ability of the different models is presented. The out-of-sample performance of the machine learning and deep learning models shows a superior forecast for SA-BiLSTM, with an RMSE of 13.128 for one-step ahead forecast. The performance of the models from the best to the least is given as SA-BiLSTM, GRU, ANN, XGBoost, LSTM, MLP, SVR, Transformer, BiLSTM, TCN and LightGBM. Four of the models have superior performance compared with nonlinear regime switching (RMSE=14.7) and six of the models have superior

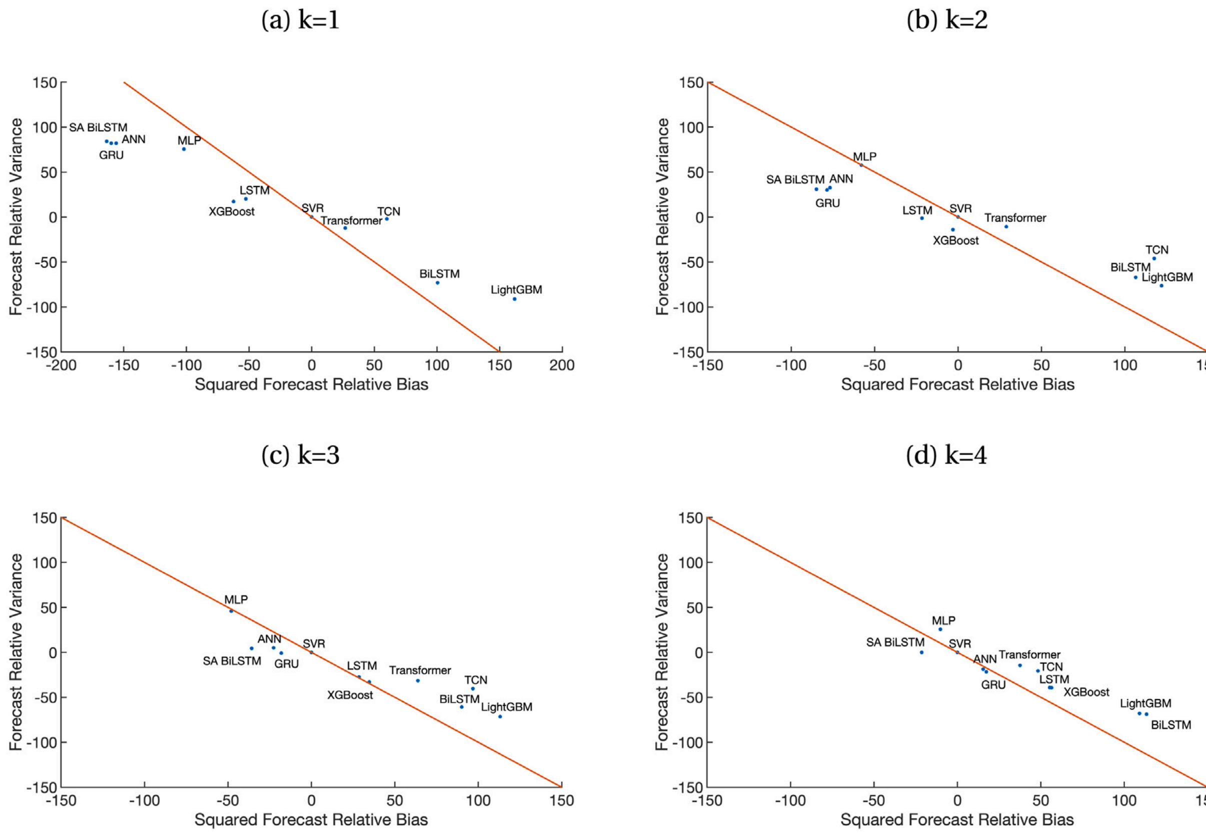

Fig. 7. Relative MSE Decomposition.
Note. The y-axis represents the relative forecast variance, while the x-axis represents the relative squared forecast bias. These values are computed as the difference between the variance (squared bias) of the considered model and the variance (squared bias) of the benchmark (SVR). The red line indicates a forecast with an equivalent mean squared error (MSE) to the benchmark. Points to the left (right) of this line represent forecasts that surpass (are surpassed by) the benchmark.

performance compared with linear (RMSE=15.4) models in Bunn et al. (2020).

Considering a two-step ahead forecast, a similar result is obtained such that the performance of the models follows this order from best to the least: SA-BiLSTM, GRU, ANN, LSTM, XGBoost, MLP, SVR, Transformer, BiLSTM, LightGBM and TCN. The proposed SA-BiLSTM has a desirable performance with an RMSE of 15.601. For a three-step ahead forecast, the models perform from the best to the least as follows: SA-BiLSTM, GRU, ANN, MLP, SVR, LSTM, XGBoost, BiLSTM, Transformer, LightGBM and TCN. The proposed SA-BiLSTM performs well with an RMSE of 16.839. A similar result is obtained for the 4-step ahead forecast. The models perform from the best to the least as follows: SA-BiLSTM, GRU, ANN, SVR, MLP, LSTM, XGBoost, Transformer, TCN, LightGBM and BiLSTM. The proposed SA-BiLSTM has an RMSE of 17.501.

5.2. Relative MSE decomposition

In this section, we investigate the bias–variance trade-off by decomposing the Mean Squared Error (MSE) of each forecasting model into two parts: the mean forecast variance and the squared mean forecast bias.

Following Lima and Meng (2017), we calculate the MSE of any forecast $\hat{Y}_{t+k|t}$ as

$$\frac{1}{P}\sum_{t+k=T_1}^{T_2}(Y_{t+k}-\hat{Y}_{t+k|t})^2$$

and the respective mean forecast variance as

$$\frac{1}{P}\sum_{t+k=T_1}^{T_2}(\hat{Y}_{t+k|t}-\frac{1}{P}\sum_{t+k=T_1}^{T_2}\hat{Y}_{t+k|t})^2$$

where P is the number of out-of-sample forecasts from T_1 to T_2. The squared mean forecast bias is computed as the difference between the MSE and the mean forecast variance.

Fig. 7 presents the relative mean forecast variance and relative squared mean forecast bias for all models and all forecasting horizons, calculated as the difference between the mean forecast variance (squared mean forecast bias) of the considered model and the mean forecast variance (squared mean forecast bias) of the machine learning benchmark model (SVR). In this way, the relative mean forecast variance and the relative mean forecast bias for the machine learning benchmark (SVR) are set to be zero. The points on the red line represent forecasts with the same MSE as the machine learning benchmark (SVR).[10] The points below (above) the red line are forecasts that outperform (are outperformed by) the machine learning benchmark (SVR).

Fig. 7 shows that models that outperform SVR with one-step-ahead forecasts are also likely to do so with multi-steps. For example, the LightGBM model is located above the red line for all forecasting periods. The LightGBM model is outperformed by the SVR benchmark model, whose forecasts are more biased and have lower variance compared with the SVR model.

Three models consistently outperform the machine learning benchmark (SVR) for all forecasting horizons; the GRU, ANN and the proposed SA-BiLSTM. The proposed SA-BiLSTM is located on the left and has the longest distance to the red line, suggesting the lowest MSE

[10] Implicit in the red line is a one-for-one tradeoff between the mean forecast variance and the mean forecast bias. While other researchers might have different preferences for this tradeoff, we follow the literature in this regards, such as Lima and Meng (2017).

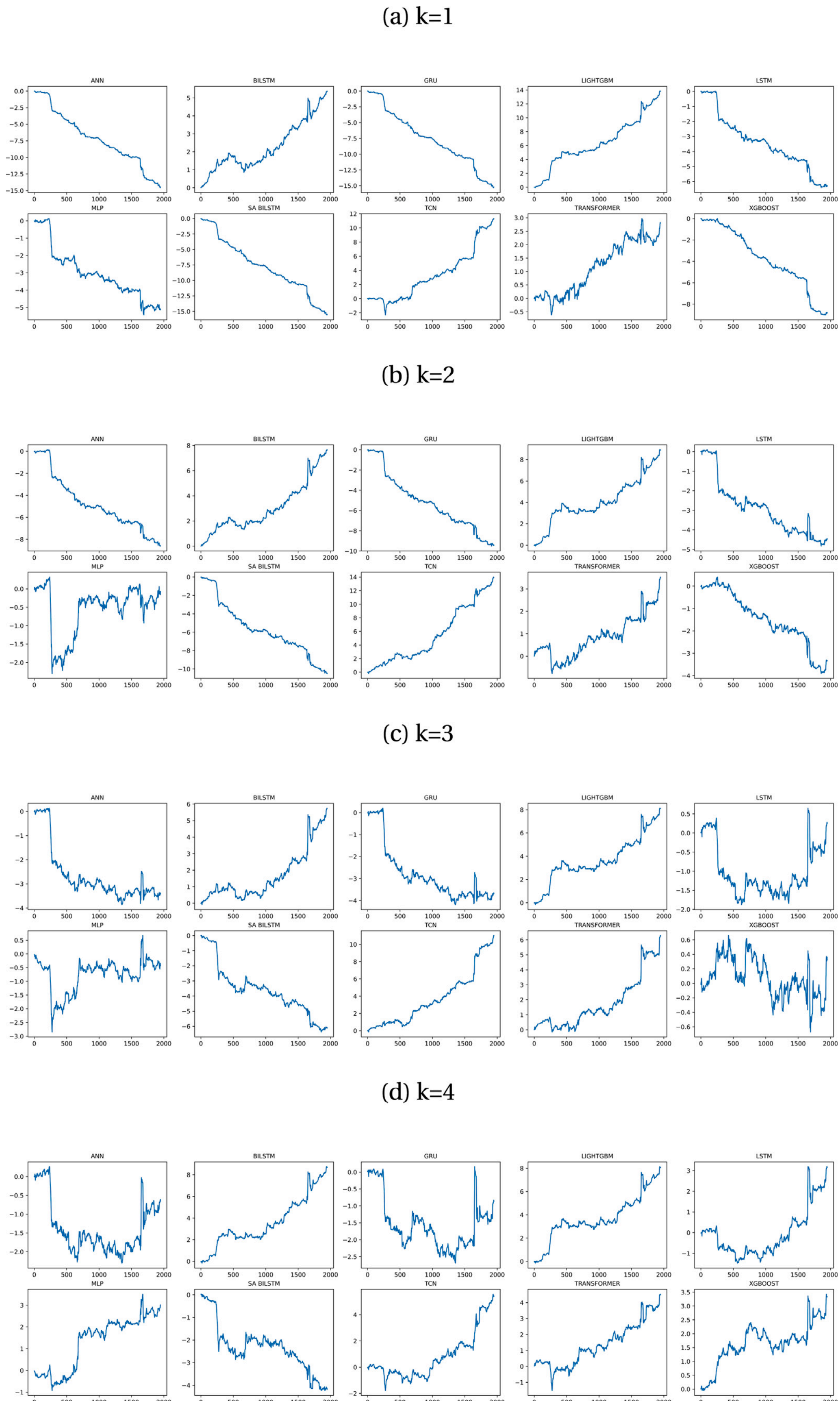

Fig. 8. Relative Cumulative Median Squared Error.

Note. This visualization illustrates the Relative Cumulative Mean Squared Errors (RCMSE) across all pseudo-out-of-sample periods. RCMSE is computed as the difference between the cumulative squared error of the model under consideration and that of the SVR benchmark. In this way, the SVR model is set to have a zero RCMSE over time.

among all forecasting horizons. It is less biased and has a higher variance than the SVR benchmark model. It is a strong signal that capturing seasonality can help generate less biased forecasts than the benchmark model but at the price of higher variance.

5.3. Relative cumulative mean squared errors

To visualize the out-of-sample forecast variance better, we plotted the Relative Cumulative Mean Squared Errors (RCMSE) for all pseudo-out-of-sample periods in Fig. 8. It is calculated as the difference between the cumulative squared error from the considered model and the cumulative squared error from the SVR benchmark. In this way, the SVR model is set to have a zero RCMSE over time. Fig. 8 reveals whether a model can consistently beat the SVR benchmark over time. A positive slope in a subplot suggests that the model is outperformed by the SVR benchmark. In contrast, a negative slope indicates that the model outperforms the SVR benchmark over time. The proposed SA-BiLSTM approach is the only model that consistently outperforms the SVR benchmark for all forecasting horizons $k = 1, 2, 3$ and 4. It is a strong signal that seasonality dramatically contributes to the consistent forecast accuracy gain over the SVR benchmark.

6. Concluding comments and future research

The importance of accurate forecasts for all participants involved in electricity markets is almost self-evident. In this paper, we study the performance of various machine learning and deep learning models in forecasting the imbalance prices in Great Britain. We compare the performance of one- to four-step-ahead forecasting models under various scenarios and measurement metrics. This allows us to demonstrate the superior forecasting accuracy of deep learning models, especially the SA-BiLSTM framework. When forecasting with the full sample (the 2016–2019 period), the SA-BiLSTM model provides a better one-step ahead predictability of imbalance prices by 11%–15% as compared with other linear, nonlinear and Bayesian models adopted in existing studies.

While we would argue our proposed model is an advancement on existing forecasting models, there are still several avenues for future research. To allow for a like-with-like comparison, we have used the same trimmed data set used in the literature. Having established the accuracy of our proposed model on this trimmed data, it would be worthwhile exploring the model's forecasting efficacy including outliers. Our proposed model, moreover, is specifically tailored to the British electricity market. Seasonality, however, is an important factor in many other contexts. This suggests that our methodology would be of use in a range of other applications. In the future we plan to investigate the robustness of our proposed model in other scenarios, in particular different energy markets.

CRediT authorship contribution statement

Sinan Deng: Writing – original draft, Formal analysis, Conceptualization. **John Inekwe:** Writing – original draft, Investigation, Conceptualization. **Vladimir Smirnov:** Writing – original draft, Formal analysis, Conceptualization. **Andrew Wait:** Writing – original draft. **Chao Wang:** Writing – review & editing, Methodology, Formal analysis, Conceptualization.

References

Bai, S., Kolter, J.Z., Koltun, V., 2018. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. arXiv preprint:1803.01271.
Borne, O., Korte, K., Perez, Y., Petit, M., Purkus, A., 2018. Barriers to entry in frequency-regulation services markets: Review of the status quo and options for improvements. Renew. Sustain. Energy Rev. 81, 605–614.
Brouwer, A.S., Broek, M.V.D., Seebregts, A., Faaij, A., 2014. Impacts of large-scale Intermittent Renewable Energy Sources on electricity systems, and how these can be modeled. Renew. Sustain. Energy Rev. 33, 443–466.
Bueno-Lorenzo, M., Moreno, M.A., Usaola, J., 2013. Analysis of the imbalance price scheme in the Spanish electricity market: A wind power test case. Energy Policy 62, 1010–1019.
Bunn, D.W., Gianfreda, A., Kermer, S., 2018. A trading-based evaluation of density forecasts in a real-time electricity market. Energies 11, 2658.
Bunn, D.W., Inekwe, J.N., MacGeehan, D., 2020. Analysis of the fundamental predictability of prices in the British balancing market. IEEE Trans. Power Syst. 36, 1309–1316.
Bunn, D.W., Kermer, S.O., 2021. Statistical arbitrage and information flow in an electricity balancing market. Energy J. 42, 19–40.
Cho, K., Merriënboer, B.V., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., Bengio, Y., 2014. Learning phrase representations using RNN encoder-decoder for statistical machine translation. arXiv preprint:1406.1078.
Friedman, J.H., 2002. Stochastic gradient boosting. Comput. Stat. Data Anal. 38, 367–378.
Garcia, M.P., Kirschen, D.S., 2006. Forecasting system imbalance volumes in competitive electricity markets. IEEE Trans. Power Syst. 21, 240–248.
Ghoddusi, H., Creamer, G.G., Rafizadeh, N., 2019. Machine learning in energy economics and finance: A review. Energy Econ. 81, 709–727.
Goodarzi, S., Perera, H.N., Bunn, D., 2019. The impact of renewable energy forecast errors on imbalance volumes and electricity spot prices. Energy Policy 134, 110827.
Hirth, L., Ziegenhagen, I., 2015. Balancing power and variable renewables: Three links. Renew. Sustain. Energy Rev. 50, 1035–1051.
Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. Neural Comput. 9 (8), 1735–1780.
Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., Liu, T.-Y., 2017. Lightgbm: A highly efficient gradient boosting decision tree. Adv. Neural Inf. Process. Syst. 30.
Klæboe, G., Eriksrud, A.L., Fleten, S.-E., 2015. Benchmarking time series based forecasting models for electricity balancing market prices. Energy Syst. 6, 43–61.
Kristiansen, T., 2007. The nordic approach to market-based provision of ancillary services. Energy Policy 35, 3681–3700.
Li, Y., Peng, T., Sun, W., Ji, C., Wang, Y., Tao, Z., Zhang, C., Nazir, M.S., 2023. A soft sensor model based on CNN-BiLSTM and IHHO algorithm for Tennessee Eastman process. Measurement 218, 113195.
Lima, L., Damien, P., Bunn, D.W., 2022. Bayesian predictive distributions for imbalance prices with time-varying factor impacts. IEEE Trans. Power Syst..
Lima, L.L., Meng, F., 2017. Out-of-sample return predictability: A quantile combination approach. J. Appl. Econometrics 32 (4), 877–895.
Lisi, F., Edoli, E., 2018. Analyzing and forecasting zonal imbalance signs in the Italian electricity market. Energy J. 39, 1–20.
Lu, Y., Suthaharan, N., 2023. Electricity price spike clustering: A zero-inflated GARX approach. Energy Econ. 124, 106834.
Lucas, A., Pegios, K., Kotsakis, E., Clarke, D., 2020. Price forecasting for the balancing energy market using machine-learning regression. Energies 13, 5420.
Marneris, I.G., Roumkos, C.G., Biskas, P.N., 2019. Towards balancing market integration: Conversion process for balancing energy offers of central-dispatch systems. IEEE Trans. Power Syst. 35, 293–303.
Mirakyan, A., Meyer-Renschhausen, M., Koch, A., 2017. Composite forecasting approach, application for next-day electricity price forecasting. Energy Econ. 66, 293–303.
Möller, C., Rachev, S.T., Fabozzi, F.J., 2011. Balancing energy strategies in electricity portfolio management. Energy Econ. 33, 2–11.
Muratori, M., Rizzoni, G., 2015. Residential demand response: Dynamic energy management and time-varying electricity pricing. IEEE Trans. Power Syst. 31, 1108–1117.
Mureddu, M., Meyer-Ortmanns, H., 2018. Extreme prices in electricity balancing markets from an approach of statistical physics. Phys. A 490, 1324–1334.
Ocker, F., Ehrhart, K.-M., 2017. The "German Paradox" in the balancing power markets. Renew. Sustain. Energy Rev. 67, 892–898.
Papadimitriou, T., Gogas, P., Stathakis, E., 2014. Forecasting energy markets using support vector machines. Energy Econ. 44, 135–142.
Poplavskaya, K., de Vries, L., 2019. Distributed energy resources and the organized balancing market: A symbiosis yet? Case of three European balancing markets. Energy Policy 126, 264–276.
Schuster, M., Paliwal, K.K., 1997. Bidirectional recurrent neural networks. IEEE Trans. Signal Process. 45, 2673–2681.
Shambora, W.E., Rossiter, R., 2007. Are there exploitable inefficiencies in the futures market for oil? Energy Econ. 29, 18–27.
Szabó, D.Z., Duck, P., Johnson, P., 2010. Optimal trading of imbalance options for power systems using an energy storage device. European J. Oper. Res. 285, 3–22.
Taylor, J.W., 2006. Density forecasting for the efficient balancing of the generation and consumption of electricity. Int. J. Forecast. 22, 707–724.
Thomaidis, N.S., Dounias, G.D., 2012. A comparison of statistical tests for the adequacy of a neural network regression model. Quant. Finance 12, 437–449.
Vandezande, L., Meeus, L., Belmans, R., Saguan, M., Glachant, J.-M., 2010. Well-functioning balancing markets: A prerequisite for wind power integration. Energy Policy 38, 3146–3154.
Vapnik, V., 1999. The Nature of Statistical Learning Theory. Springer Science & Business Media.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I., 2017. Attention is all you need. Adv. Neural Inf. Process. Syst.

Wu, Z., Zhou, M., Zhang, T., Li, G., Zhang, Y., Liu, X., 2020. Imbalance settlement evaluation for China's balancing market design via an agent-based model with a multiple criteria decision analysis method. Energy Policy 139, 111297.

Xiong, J., Peng, T., Tao, Z., Zhang, C., Song, S., Nazir, M.S., 2023. A dual-scale deep learning model based on ELM-BiLSTM and improved reptile search algorithm for wind power prediction. Energy 266, 126419.

Zhang, C., Ma, H., Hua, L., Sun, W., Nazir, M.S., Peng, T., 2022. An evolutionary deep learning model based on TVFEMD, improved sine cosine algorithm, CNN and BiLSTM for wind speed prediction. Energy 254, 124250.