Applied Energy 353 (2024) 122059

Contents lists available at ScienceDirect

Applied Energy

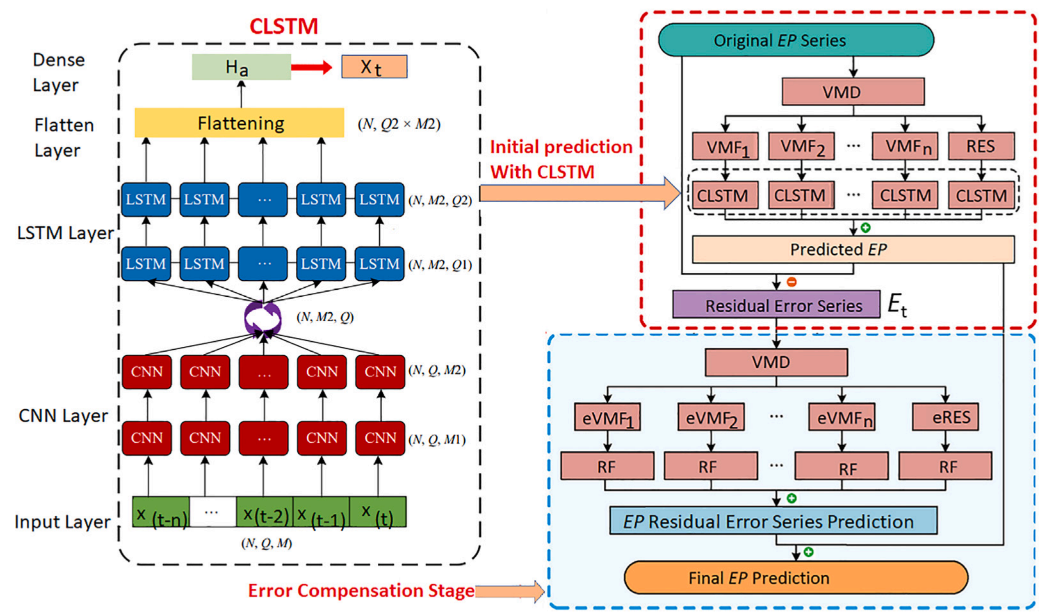
journal homepage: www.elsevier.com/locate/apenergy

Two-step deep learning framework with error compensation technique for short-term, half-hourly electricity price forecasting

Sujan Ghimire [a], Ravinesh C. Deo [a,b], David Casillas-Pérez [c,*], Sancho Salcedo-Sanz [d,a]

[a] School of Mathematics, Physics, and Computing, University of Southern Queensland, Springfield, QLD, 4300, Australia
[b] Centre for Applied Climate Sciences, University of Southern Queensland, Toowoomba, QLD, 4500, Australia
[c] Department of Signal Processing and Communications, Universidad Rey Juan Carlos, Fuenlabrada, 28942, Madrid, Spain
[d] Department of Signal Processing and Communications, Universidad de Alcalá, Alcalá de Henares, 28805, Madrid, Spain

GRAPHICAL ABSTRACT



ARTICLE INFO

Dataset link: https://www.aemo.com.au/

Keywords:
Electricity price prediction
Deep learning
Error compensations
Variational mode decomposition
Hybrid model

ABSTRACT

Prediction of electricity price is crucial for national electricity markets supporting sale prices, bidding strategies, electricity dispatch, control and market volatility management. High volatility, non-stationarity and multi-seasonality of electricity prices make it significantly challenging to estimate its future trend, especially over near real-time forecast horizons. An error compensation strategy that integrates Long Short-Term Memory (LSTM) network, Convolution Neural Network (CNN) and the Variational Mode Decomposition (VMD) algorithm is proposed to predict the half-hourly step electricity prices. A prediction model incorporating VMD and CLSTM is first used to obtain an initial prediction. To improve its predictive accuracy, a novel error compensation framework, which is built using the VMD and a Random Forest Regression (RF) algorithm, is

* Corresponding author.
E-mail addresses: sujan.ghimire@usq.edu.au (S. Ghimire), ravinesh.deo@usq.edu.au (R.C. Deo), david.casillas@urjc.es (D. Casillas-Pérez), sancho.salcedo@uah.es (S. Salcedo-Sanz).

https://doi.org/10.1016/j.apenergy.2023.122059
Received 25 July 2023; Received in revised form 5 September 2023; Accepted 30 September 2023
Available online 19 October 2023

also used. The proposed VMD-CLSTM-VMD-ERCRF model is evaluated using electricity prices from Queensland, Australia. The results reveal highly accurate predictive performance for all datasets considered, including the winter, autumn, spring, summer, and yearly predictions. As compared with a predictive model without error compensation (i.e., the VMD-CLSTM model), the proposed VMD-CLSTM-VMD-ERCRF model outperforms the benchmark models. For winter, autumn, spring, summer, and yearly predictions, the average Legates and McCabe Index is seen to increase by 15.97%, 16.31%, 20.23%, 10.24%, and 14.03%, respectively, relative to the benchmark models. According to the tests performed on independent datasets, the proposed VMD-CLSTM-VMD-ERCRF model can be a practical stratagem useful for short-term, half-hourly electricity price forecasting. Therefore the research outcomes demonstrate that the proposed error compensation framework is an effective decision-support tool for improving the predictive accuracy of electricity price. It could be of practical value to energy companies, energy policymakers and national electricity market operators to develop their insight analysis, electricity distribution and market optimization strategies.

1. Introduction

Renewable energy sources such as solar and wind are becoming prevalent in power systems supporting national electricity markets, making it more difficult to balance electricity supply and demand, costs and affordability. This calls for electricity demand response strategies that can analyse consumer demand patterns. Many countries are already employing dynamic pricing, which includes periodically adjusting electricity prices (*EP*). Consumers, especially prosumers who produce their own electricity, must accurately predict changes in *EP*. Credible *EP* predictions enable prosumers to optimize their energy use, receive incentives on electricity bills, and contribute to the stability of the regional and national energy grid [1]. As such, in a highly competitive and rapidly evolving national electricity market, it is imperative for market participants to anticipate *EP* shifts and dynamic market-driven movements accurately.

Predicting electricity prices is a complex task due to the unique characteristics of *EP*, especially over short periods. The most challenging features of *EP* include seasonality recorded at many frequencies, jumps in both directions (positive and negative) and high volatility on a daily or hourly basis. This challenging behaviour of *EP* has attracted the attention of many research scholars [2].

Statistical, computational and hybrid models are the three main branches for the development of *EP*. Traditional statistical such as Regression Model, Transfer Function, Exponential Smoothing, Autoregressive Moving Average, Autoregressive Integrated Moving Average, Autoregressive Fractionally Integrated Moving Average and Generalized Autoregressive Conditional Heteroskedasticity with their improved versions such as Autoregressive Moving Average Exogenous, are utilized for *EP* prediction. These simple methods take into account time-based relationship of data, making them appropriate for predicting the *EP* series with minor fluctuations and with low-frequency changes (relying on the high stability of data patterns) [3]. However, electricity demand and price data that have a high degree of randomness and intermittent patterns due to consumer's purchase and use of electricity with the *EP* time series data generally comprising of complex features e.g., high and low frequencies, volatility, variable means and variances, and a high proportion of unusual prices [4,5]. Consequently, conventional methods are limited in their accuracy in predicting *EP*.

Computational intelligence models, facilitated by Artificial Intelligence (AI) techniques, are capable of extracting complex nonlinear features in electricity price datasets. These methods do not require meeting specific statistical assumptions, and have a higher accuracy for predicting nonlinear time series data [6]. Common models used for *EP* prediction are presented in Table 1. Broadly categorized into Machine Learning (ML) and Deep Learning (DL) models, these methods have a better prediction result compared to traditional methods due to their robustness and nonlinear mapping capabilities. However, they still fall short in exploring the internal time dynamics of time-series data [7].

DL methods specifically designed for sequence-based models include Recurrent Neural Networks (RNN) [8] and their variations including Long Short-Term Memory (LSTM) [9–11], Deep Belief Network (DBN) [12], Auto-Encoder (AE) [13,14], Convolution Neural Network (CNN) [15,16] and Gated Recurrent Unit (GRU) [17]. These are particularly effective in handling continuous sequences, making them highly applicable for *EP* prediction. Lago et al. [18] compared several DL models like DNN, GRU, and LSTM, and ML models like RF and RBFNN as well as the ARIMA statistical model to predict spot *EP* using European power exchange (EPEX) Belgium datasets. They found that DL models were more accurate in terms of Symmetric Mean absolute Percentage Error. Specifically, for daily *EP* prediction, this error was 12.34%, 13.04%, and 13.06% for DNN, GRU, and LSTM, respectively, while it was 14.77%, 15.39%, and 19.32% for RBFNN, RF, and ARIMA, respectively.

Gokgoz and Filiz [19] compared ANN with DNN models for *EP* prediction to show that the DNN model, utilizing data from 10–50 days prior, performed better than ANN with a Mean Absolute Error (*MAE*;USD/MWh) of 0.346. Wang et al. [20] suggested a DL model for short-term hourly prediction of *EP* using a Stacked Denoising Auto Encoder model (SDA). Their SDA model generated a *MAPE* of 4.45%, lower than that of other models like ANN, Multivariate Adaptive Regression Splines (MARS), SVM, and Least Absolute Shrinkage and Selection Operator (LASSO), which generated MAPE values of 5.52%, 5.79%, 6.22%, and 7.56%, respectively.

In spite of the success of AI models, the complexity of *EP* makes it difficult to achieve optimal prediction using a single model [21], and especially when the hyperparameters are not well-tuned [22]. Hybrid models have therefore become a prevailing method for *EP* predictions [23] as they integrate several AI-based models with model's input data decomposition for improved performance. For example, a hybrid model may consists of two variants: the first as a statistical model and anther as a ML model such as ARFIMA-ANN [24], ARMA-ELM [25], and ARMAX-LSSVM [26]. However, this type of architecture has known limitations such as greater model complexity, computational costs and reduced model interpretability.

Many researchers are now moving towards integrating several ML models for superior accuracy and flexibility through methods like ELM-ANN [57] and LSTM-ANN [58] approaches. The combination of DL models like CNN-LSTM [59], has demonstrated the most promising result. In hybrid CNN-LSTM model, the CNN layer is capable of extracting features among several variables that influence *EP* [60]. Kuo et al. [61] introduced a short-term *EP* prediction model utilizing a CNN-LSTM hybrid neural network that takes into account the real-time *EP*. The hybrid CNN-LSTM model outperformed other models such as SVM, RF, MLP, CNN, and LSTM in terms of MAE. Specifically, the *MAE* for CNN-LSTM was 8.84, which is lower than LSTM (9.82), CNN (9.80), MLP (9.86), RF (9.20), and SVM (28.98). Similarly, Heidarpanah et al. [62] employed a CNN-LSTM model to predict *EP* in Iran's electricity market. The CNN-LSTM model was compared to Multivariate Linear Regression (MLR), SVM, ANN, ANFIS, and ANN-Genetic Algorithm models. In Iran's electricity market, the hybrid CNN-LSTM model was found to be the most robust. In addition, the ANN, ANN-GA, and ANFIS models showed acceptable results. However, MLR and SVM models failed to account for *EP* time-series' sinusoidal and fluctuating nature.

Table 1
Selected research on electricity price forecasting based on machine learning algorithms.

Reference	Year	Machine learning algorithm	Country	Model inputs
[27]	2023	Stacked AEs	PJM, USA	Historical time-series of *EP*
[28]	2023	ELM	PJM,USA; AEM, Australia and OEM, Canada	Historical time-series of *EP*
[29]	2023	ARIFMA, GARCH	Italy; Belgium	Historical time-series of *EP*, Load, Solar and wind generation, hydro, biomass and waste generation
[30]	2023	ANN	Russia	Historical time-series of *EP*
[31]	2023	CNN-BiLSTM-AR	Nord Pool energy market, EU	Historical time-series of *EP*
[32]	2023	STL-TCN-NBEATS	Spain	Historical time-series of *EP*, electricity consumption, power generation, and weather data
[33]	2023	SSA-NBEATS	Shanxi, China	Historical time-series of *EP*
[34]	2023	LR-CatBoost	Nord Pool energy market, EU	Historical time-series of *EP*
[35]	2022	SSA-DELM	Denmark	Historical time-series of *EP*, Load, Solar and wind generation
[36]	2022	GPR, ANN	UK, Germany, Denmark and Sweden	Historical time-series of *EP*, Load, Solar and wind generation, hydro, biomass, geothermal and waste generation, Fossil fuel prices and policy instruments
[37]	2022	IDPSO-VMD-XGB	Greece	Historical time-series of *EP*, Temperature and Humidity
[38]	2022	ERC–DNN	Nord Pool energy market, EU	Historical time-series of *EP*, Solar and wind generation
[39]	2022	NARMAX	Irish Integrated Single Electricity Market, Ireland	Historical time-series of *EP*, Load, Temperature, C02 emission e.t.c
[40]	2022	ILRCN	Texas, USA	Historical time-series of *EP* and Load,
[41]	2021	LSTM	Nord Pool	Historical time-series of *EP*
[42]	2021	GRU	PJM, USA	Historical time-series of *EP*
[43]	2019	CNN	Ireland	Historical time-series of *EP*
[44]	2019	BDL	Italy; Belgium	Historical time-series of *EP*
[45]	2018	RNN	Turkey	Historical time-series of *EP*
[46]	2016	DBN	Macedonia	Historical time-series of *EP*
[47]	2015	GRNN	Spain	Historical time-series of *EP* and load
[48]	2015	RF	German	Historical time-series of *EP*
[49]	2014	LSSVM	Nord Pool energy market, EU	Historical time-series of *EP*
[50]	2010	RBFNN	PJM, USA	Historical time-series of *EP*
[51]	2010	PNN	PJM, USA; QLD, Australia	Historical time-series of *EP*
[52]	2010	ANFIS	Spain	Historical time-series of *EP*
[53]	2010	SVM	California, USA	Historical time-series of *EP*
[54]	2008	WNN	Spain; PJM, USA	Historical time-series of *EP*
[55]	2004	ANN	California, USA	Historical time-series of *EP* and load
[56]	1999	BPNN	Victoria, Australia	Historical time-series of *EP*

The present study develops a two-step framework with error compensation strategy for short-term, half-hourly electricity price forecasting using data decomposition algorithms. In general, the *EP* time-series datasets are relatively complex with intertwined features, short-term cyclic changes in price, long-term trends, and erratic spikes reflecting varying degree of consumption and productions. Attaining accurate *EP* predictions using a simple hybrid model is challenging without first decomposing and revealing the data features. One new approach that this paper presents, involves a data decomposition pre-processing technique to alleviate the impact of chaotic features. Different decomposition methods have so far been integrated into hybrid models for *EP* prediction, which commonly transform the original *EP* time-series sequences into sub-series that exhibit more stable variance and fewer outliers. These sub-series can be combined with DL algorithms to create hybrid models to attain accurate prediction results. Examples of include Wavelet Transform (WT) [63], Empirical Mode Decomposition (EMD) [10,64], Variational Mode Decomposition (VMD) [65], Singular Spectrum Analysis (SSA) [66], Ensemble Empirical Mode Decomposition (EEMD) [67], Complete Ensemble Empirical Mode Decomposition (CEEMD) [68], and Improved Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (ICEEMDAN) [69,70]. Qiao et al. [71] used WT, SAE, and LSTM models to generate price predictions for United State electricity markets. Although SAE-LSTM was found to have better prediction accuracy, the WT-SAE-LSTM model was deemed to have more practical value. Conejo et al. [72] proposed a hybrid model based on WT and the ARIMA model to predict *EP* in Spain's market. Meanwhile, Hannah Jessie Rani and Aruldoss Albert Victoire [65] predicted multi-step *EP* in a power system through an improved VMD method and ANN. Huang et al. [73] focused on short-term *EP* and used a hybrid model incorporating VMD, CNN and GRU. Wang et al. [74] analysed *EP* in the Australian and French markets, constructing a hybrid model based on fast EEMD, VMD, and a back propagation NN optimized using the Firefly Algorithm (FA).

Chang et al. [75] developed a hybrid model that combined WT and LSTM and evaluated its performance using datasets from New South Wales, Australia, and France. As in previous studies, WT was used to decompose the data. The *EP* time-series was initially broken down into several component series with minor variances. The decomposed time-series were then separately trained and predicted using LSTM, and the predicted values were summed to generate the final prediction. With WT, the variance of the time-series data became more stable, allowing LSTM to capture fluctuations in *EP* more accurately and considerably improve the prediction accuracy compared to a model that combined ARIMA and ANN models [76,77].

In accordance with literature, we note that data decomposition models show that, while wavelet transform (WT) is unable to adapt to different scenarios and extract detailed information [78], EMD and its variants (EEMD, CEEMD) are relatively susceptible to noise and sampling errors. Conversely, the VMD method is a superior technique that effectively overcomes the limitations of other models. Also, although CNN-LSTM have been widely used, there is insufficient coverage of the impact of error compensation on state-of-the-art CNN-LSTM. While many studies use hyperparameter optimization and feature selection to fine-tune models and achieve lower error metrics, few have employed decomposition techniques to refine and improve predictions. Consequently, the potential use of pre-processing of *EP* time-series and an error estimation module for benchmarks utilizing the CNN-LSTM model as an additional tuning tool remains an open question.

The main contribution and scientific novelty of this paper is to propose for the first time a hybrid predictive modelling approach that involves a two-stage decomposition-based Error Compensation (ERC) model. The proposed VMD-CLSTM-VMD-ERCRF model incorporates a number of model input data processing stages that can enhance its accuracy relative to a standalone model. First, the original *EP* series are decomposed into sub-series and a residual series using VMD algorithm. An CLSTM network is applied to predict each sub-series. Next, an error series is constructed by comparing the predicted sub-series with the original observation value. This error series is further decomposed using VMD to obtain sub-series, which are predicted using an Random Forest (RF) network. The predicted error series is then used to compensate the prediction result of the original series, resulting in the final predicted half-hourly*EP* series. The proposed VMD-CLSTM-VMD-ERCRF model can therefore consider the relatively complex, antecedent *EP* time series to predict the future value.

As a commitment to demonstrate significant improvements in *EP* prediction, this study aims to make primary contributions to accurate electricity price forecasting. To accomplish this, we investigate the stability of predicted sequences after the ERC stage and provide further insight into the suitability of Error Compensation modules for future integration with modern benchmark models. Baseline VMD-CLSTM, LSTM, DNN, XGB, and RF model were optimized using the Bayesian hyperparameter optimization method, and resulting error metrics were compared to the proposed approach on the *EP* time-series dataset of Queensland, Australia. This framework, as shown later in Fig. 6, has an error compensation stage which is activated after the CLSTM model's initial prediction, while the error compensation stage adopts a residual error series prediction step with an RF method to improve the final outcome of the VMD-CLSTM-VMD-ERCRF model. The hybrid VMD-CLSTM-VMD-ERCRF model therefore presented accurately captures nonlinear characteristics of *EP*, resulting in accurate prediction, providing a new perspective on *EP* prediction.

2. Materials and method

2.1. Description of electricity price datasets

To ascertain the effectiveness and efficiency of the proposed VMD-CLSTM-VMD-ERCRF model for half-hourly electricity price predictions, rigorous tests on the Australian National Electricity Market (ANEM) datasets, which encompasses five different regional market jurisdictions, namely Queensland, New South Wales, Victoria, South Australia, and Tasmania, were performed.

The Australian Energy Market Operator (AEMO) (https://www.aemo.com.au/) manages the entire power system. In the ANEM space, electricity trading is conducted through half-hourly trading intervals. Therefore the power generators are required to submit their price offers, which are dispatched with corresponding dispatch prices every five minutes. The market clearing price is determined by averaging the six consecutive 5-min dispatch prices for each half-hour interval, based on the bids and offers of scheduled generators and consumers. This process results in a separate spot price being determined for each of the five regions within AEMO.

This study has employed half-hourly electricity price sequences obtained from the AEMO for the Queensland region to evaluate the predictive performance of the proposed VMD-CLSTM-VMD-ERCRF hybrid model. The historical electricity price dataset used in this study consists of 153,793 half-hourly electricity prices recorded over a period of 3286 days from January 1, 2014 to October 10, 2022. To ensure the accuracy of the prediction model, it is necessary to limit the electricity prices within a certain range to account for potential electricity price spikes caused by factors such as power failures, transmission line maintenance, and extreme weather conditions [79]. The range set for this study was $[0, 1000]$ with $\mathrm{AUD}\ 1000/\mathrm{MWh}$ (where AUD represents the Australian Dollar) assigned to electricity prices exceeding $\mathrm{AUD}\ 1000/\mathrm{MWh}$ and $\mathrm{AUD}\ 0/\mathrm{MWh}$ assigned to electricity prices below $\mathrm{AUD}\ 0/\mathrm{MWh}$ (negative electricity prices are allowed in the Australian Electricity Market).

Out of the total dataset, there were 581 instances of electricity prices exceeding $\mathrm{AUD}\ 1000/\mathrm{MWh}$ and 2343 instances of electricity prices below $\mathrm{AUD}\ 0/\mathrm{MWh}$ during this period, which represented only 1.90% of the total data and had a limited impact on the prediction model [80]. Taking into account the distinctive seasonal variations observed within the *EP* time series derived from AEM, it becomes crucial to evaluate how different seasons influence the accuracy and stability of the proposed model. In response to this consideration, the*EP* dataset is partitioned into five segments, as outlined in Table 2. Moreover, Table 2 furnishes statistical insights for these diverse datasets, encompassing mean values, maximum (Max) values, minimum (Min) values, standard deviations (Std), skewness ($Skew$), and kurtosis ($Kurt$). This comprehensive set of statistics facilitates an in-depth analysis of the data. Notably, all five datasets ($DS1$, $DS2$, $DS3$, $DS4$, and $DS5$) exhibit kurtosis values surpassing 3, indicating that the electricity price distribution exhibits fat tails, signifying an increased likelihood of extreme values. Additionally, the EP series demonstrates a skewness exceeding 1, suggesting a significant skew in the distribution. Furthermore, for each dataset, a 20% portion of the training data is allocated for validation purposes. As an example, for $DS1$, the total data-points amount to 149,158, with 116,013 designated for training, and 29,033 and 4142 allocated for validation and testing, respectively (Table 2).

Fig. 1 displays hourly variations in electricity prices across four seasons in QLD (Summer, Autumn, Winter, Spring) demonstrating that the electricity prices are lower during Summer and Spring compared to the Winter and Autumn seasons. During all seasons, from 3:00 PM to 9:00 PM, electricity prices are higher than other times of the day.

To predict half-hourly *EP*, the proposed VMD-CLSTM-VMD-ERCRF hybrid model combines VMD-based frequency decomposition technique with a hybrid Deep Learning method (CNN-LSTM). In order to explain the proposed methods, this paper first presents the theoretical background of the VMD, followed by a discussion of the details of the proposed hybrid model. To keep this section concise, we will not provide a detailed explanation of the theoretical foundations of CNN, LSTM, XGB, RF, and DNN, as there are already numerous resources [9,10,17,81,81–83] available on these models.

2.2. Variational mode decomposition

To develop the proposed VMD-CLSTM-VMD-ERCRF model, we adopt the quasi-orthogonal decomposition technique (Variational Mode Decomposition, VMD) proposed by [84]. See Appendix B.1 for related theory of this method.

This was an adaptive method that uses a mathematically structured approach. VMD decomposes input signals into narrow-band and stationary signals (IMFs), while also allowing the possibility of reconstructing the original input signal. The VMD method employs Wiener Filtering (WF), Hilbert Transform and Heterodyne Demodulation (HHT), along with an Alternate Direction Multiplication Method (ADMM), to obtain decomposition modes. Decomposed modes are concentrated around specific central frequencies.

To create the bandwidth of a decomposed mode, different approaches can be utilized: (i) the Hilbert Transform is used to estimate the one-sided frequency spectrum of real signals using analytic representations, (ii) the base-band frequency spectrum is shifted to the estimated base-band frequency using modulation properties, and (iii) the Gaussian smoothness is used to estimate the bandwidth of the demodulated signal.

Before applying VMD, the parameters K, α, τ and ϵ (see Appendix B) are defined where the number of mode components (K) is considered most important parameter. A large K can result in intermittent decomposition results without clear patterns and reduce computational efficiency, while a small value of K can lead to insufficient decomposition accuracy, with multiple frequency components appearing simultaneously in the same mode component.

Table 2
Description of electricity price data in five experimental cases. Note *Max*, *min* and *std* refers to the Maximum, Minimum and Standard Deviation value of *EP* in *AUD/MWh* respectively. *Skew* refers to the Skewness, *Kurt* refers to the Kurtosis. *N* refers to the dataset size. Validation datasize was set as 0.2 i.e. 20% of data is used for validation.

Data set	Training samples						Validation samples	Testing samples					
Winter season prediction (DS1)	1 January 2014 to 31 May 2022							1 June 2022 to 31 August 2022					
	Max	Min	Std	Skew	Kurt	N	N	Max	Min	Std	Skew	Kurt	N
	1000.00	0.00	51.46	4.44	38.72	116,013	29,003	999.55	0.00	166.59	0.67	0.58	4142
Autumn season prediction (DS2)	1 January 2014 to 28 February 2022							1 March 2022 to 31 May 2022					
	Max	Min	Std	Skew	Kurt	N	N	Max	Min	Std	Skew	Kurt	N
	1000.00	0.00	42.16	4.73	56.50	112,530	28,132	1000.00	0.00	119.44	1.14	3.08	4353
Spring season prediction (DS3)	1 January 2014 to 31 August 2021							1 September 2021 to 31 November 2021					
	Max	Min	Std	Skew	Kurt	N	N	Max	Min	Std	Skew	Kurt	N
	973.32	0.00	39.71	4.09	40.84	106,082	26,520	1000.00	0.00	63.11	6.14	68.93	3805
Summer season prediction (DS4)	1 January 2014 to 30 November 2021							1 December 2021 to 28 February 2022					
	Max	Min	Std	Skew	Kurt	N	N	Max	Min	Std	Skew	Kurt	N
	1000.00	0.00	40.61	4.40	48.40	109,126	27,281	1000.00	0.00	67.75	6.09	66.89	4255
Year 2022 prediction (DS5)	1 January 2014 to 31 December 2021							1 January 2022 to 10 October 2022					
	Max	Min	Std	Skew	Kurt	N	N	Max	Min	Std	Skew	Kurt	N
	1000.00	0.00	41.15	4.62	53.86	111,011	27,752	1000.00	0.00	138.21	1.37	2.78	13,005

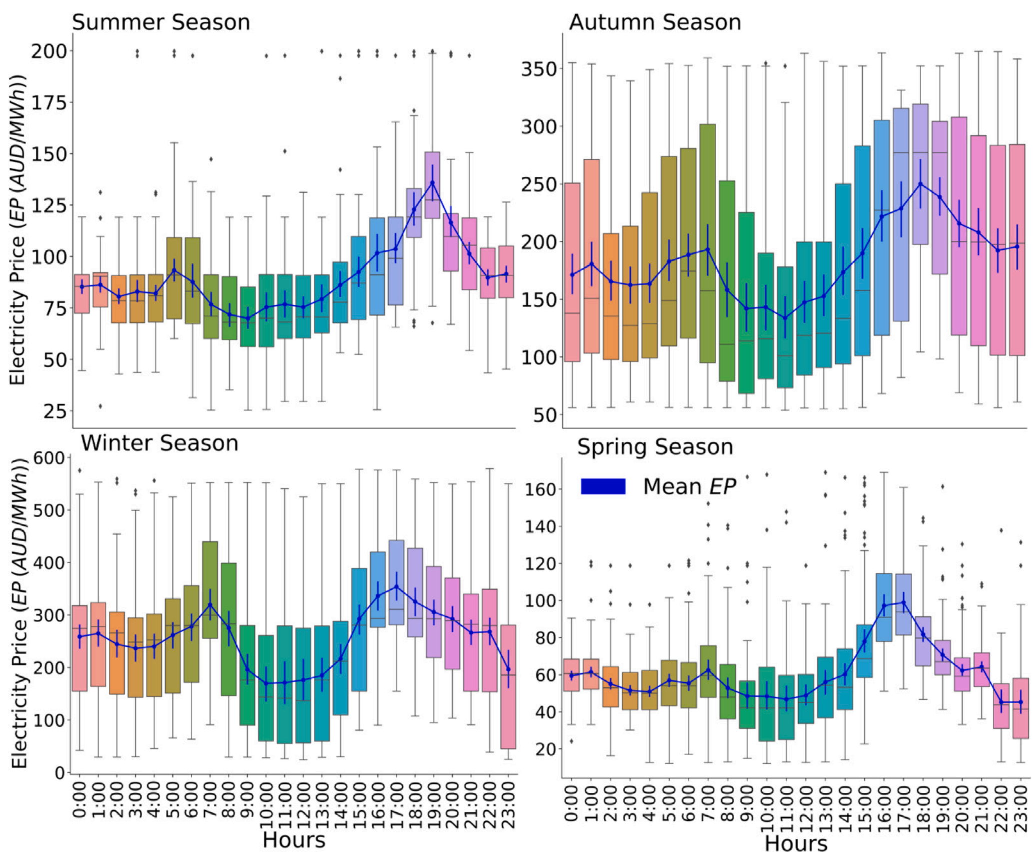


Fig. 1. The seasonal hourly distribution of *EP* in 2022. Continuous line at the centre represents the median.

In this study, VMD decomposition is performed with values of K ranging from 1 to 40, and the optimal value of K is chosen based on the Central Frequency Ratio (CFR), with the optimal value being determined when CFR tends to be stable. For the remaining input parameters, the quadratic penalty term is set to $\alpha = 2000$, the noise tolerance is set to $\tau = 0$, and the convergence criterion is set to $\epsilon = 10^{-7}$.

2.3. Hybrid convolutional and long short-term memory model

To predict *EP* using CNN-LSTM, a CNN and LSTM model has been connected in a series as a deep learning hybrid method to extract complex features from *EP* time-series and storing, complex irregular trends. The CNN layer comprises a convolution and pooling layer, from which the spatial characteristics of an *EP* time- series variable are extracted and transmitted to the LSTM layer. The LSTM layer models irregular time information using the transmitted spatial features. The general model design for predicting *EP* through CNN-LSTM model is provided in Fig. B.15 in Appendix B.3.

The LSTM, a lower layer of CNN-LSTM, stores time information regarding critical *EP* characteristics that have been extracted through CNN. The LSTM model provides a solution by retaining long-term

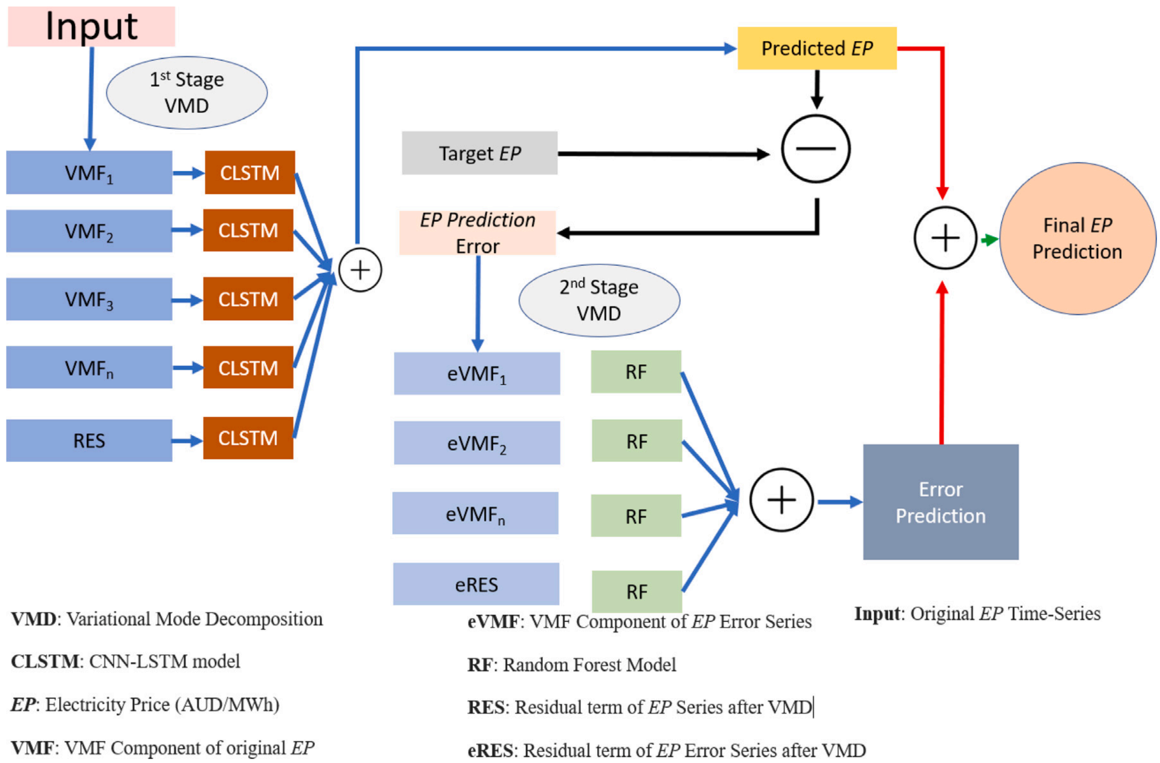


Fig. 2. The proposed VMD-CLSTM-VMD-ERCRF hybrid model, which comprises of a double decomposition method to split the electricity price data features, the CNN-LSTM network and the Random Forest models for prediction error compensations.

memory via memory units that can update the previous hidden state, thus enabling the comprehension of temporal relationships in long-term sequences. Output values from the previous CNN layer are transmitted to the gate units. The LSTM network is well-suited for *EP* prediction, as it addresses issues such as explosive and vanishing gradient problems that can arise when learning traditional RNN. The gate units are a mechanism for determining the state of each individual memory cell through multiplication operations, consisting of input, output, and forget gate units, depending on the function. For all related theory on LSTM, see Appendix B.2.

The final layer of CNN-LSTM model comprises fully connected layers that can be utilized to predict the electricity prices for a given time period (see Fig. B.15). The LSTM unit's output is flattened to a feature vector $h^l = \{h_1,\ h_2,\ \dots\ h_l\}$, where l is the number of units in LSTM. The output of the LSTM is used as input to the fully connected layer, see Eq. (1):

$$d_i^l = \sum_j w_{ji}^{l-1}(\sigma\left(h_i^{l-1}\right) + b_i^{l-1}), \tag{1}$$

where σ is a non-linear activation function, w is the weight of the ith node for layer $l-1$ and the jth node for layer l, and b_i^{l-1} represents a bias.

2.4. Proposed structure of VMD-CLSTM-VMD-ERCRF model for electricity price prediction

The primary focus of this study was to create a hybrid predictive model that can utilize a combination of double decomposition and deep learning i.e., CNN-LSTM methods, along with an error compensation module that can refine the final predictions to an higher degree of accuracy in electricity price predictions.

Fig. 2 shows a schematic of the proposed VMD-CLSTM-VMD-ERCRF model, which also outlines our methodological contributions made in respect to the error corrections for improved electricity price predictions. Firstly, the time-series data were split into the training and testing sets, and VMD techniques were applied to decompose each set of variables into VMF components and residuals (RES). This stage was very important considering the highly stochastic nature of electricity price datasets. The splitting of these data ensured that the trends, periodicity, jumps, and other rapid fluctuations were clarified visually in order to build a robust model. Secondly, the CNN-LSTM model was used to train each of the VMF and RES component separately and obtain their prediction results. Thirdly, the residual prediction error for each half-hourly sequence was calculated from the training set and decomposed using VMD again to obtain $eVMF$ and $eRES$ components. The use of residual prediction errors into VMD-CLSTM-VMD-ERCRF model is a major contribution of this study.

Using historical errors for each sub-sequence, the Random Forest (RF) model was used to estimate the error values of the next half-hour electricity price. The final price prediction was derived by adding the estimated error and the price prediction of the CNN-LSTM. The study employing the proposed VMD-CLSTM-VMD-ERCRF model therefore featured several experiments performed on different training scenarios for the interpretation and further analysis of the error compensation processes to optimize it capability to predict *EP*.

2.5. Development of VMD-CLSTM-VMD-ERCRF model

2.5.1. Variational mode decomposition

This study has utilized Keras as an open-source Python library [85] on Intel Core i7-6700k CPU with a clock speed of 4.00 GHz and 32 GB memory. The *EP* data were divided into training and testing sets as per Table 2. The training and testing sets were decomposed into K Variational Mode Functions (VMFs) using Variational Mode Decomposition (VMD) algorithm. To determine the appropriate number of VMFs (K), the Central Frequency Ratio (CFR) method was utilized, which monitors the pattern of the highest centre frequencies. As the number of mode components increases, the maximum centre frequency of each component gradually increases until reaching a steady state, at which point K can be calculated. The VMD algorithm's three other

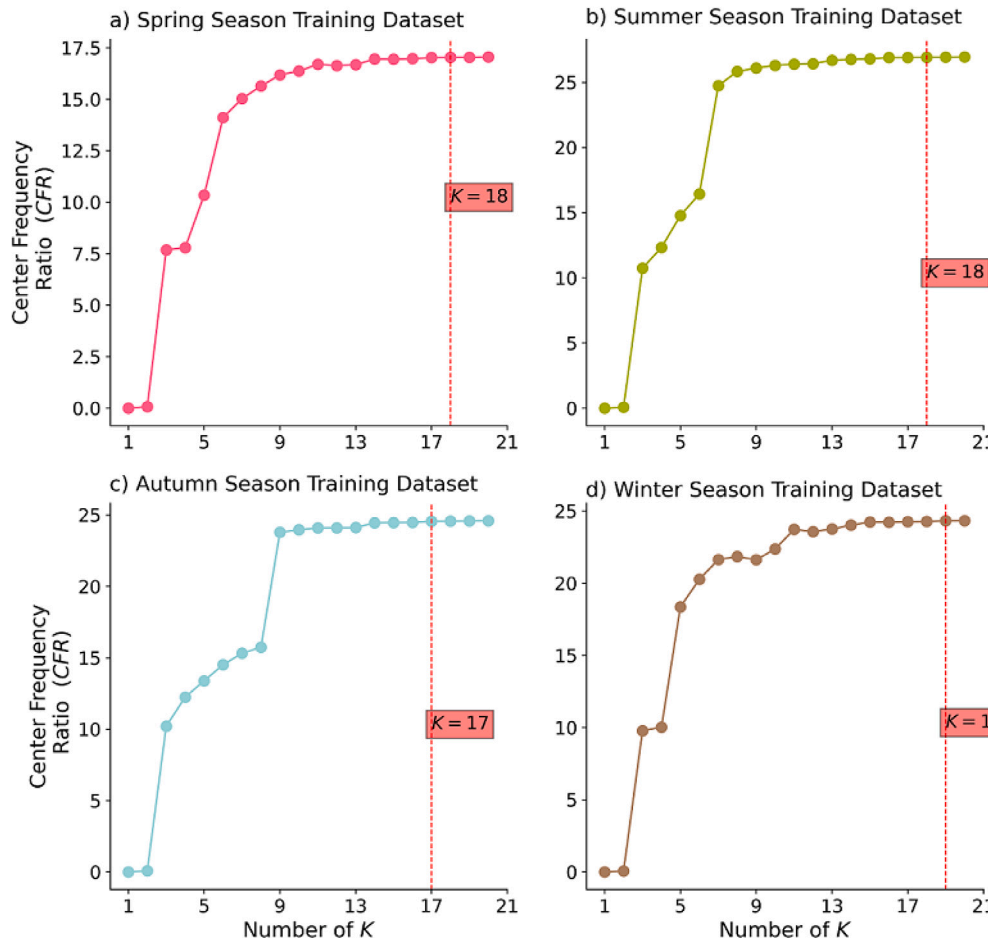


Fig. 3. Centre Frequency Ratio (CFR) corresponding to different mode number, K.

parameters, namely α (the quadratic penalty term), τ (the noise tolerance), and ε (the convergence criterion), were set to 2000, 0, and 10^{-7}, respectively.

Fig. 3 displays the centre frequency ratio for four different training datasets, and the value of K was selected as 18 for the *DS1* dataset (Spring season training dataset). Similarly, K was determined to be 18, 17, 19 and 19 for the *DS2*, *DS3*, *DS4* and *DS5* training datasets, respectively, using the CFR method. Fig. 4 shows the trends in VMF components obtained after applying the VMD to *EP* (*DS5*). Fig. 4 also displays the residual component (RES), calculated by subtracting the sum of each VMF component from original *EP* time-series. Note that for clarity, only first 500 data points for shown.

2.5.2. Pre-processing electricity price data and model variable selection

When dealing with the decomposed electricity price data, the resulting VMF signals may have a different range of values, making a comparison among them relatively challenging. To address this issue, we adopted normalization methods to ensure that the standardized data values are comparable as model inputs to prevent an ill-conditioned model and ensure a stable convergence of the weights and hyperparameters. A min–max scaling method was adopted as follows:

$$VMF_{n,norm} = \frac{VMF_n - VMF_{n,min}}{VMF_{n,max} - VMF_{n,min}}. \tag{2}$$

Here, $VMF_{n,norm}$ = normalized values, VMF_n = initial value of VMF, $VMF_{n,max}$ represents the maximum of the whole VMF and $VMF_{n,min}$ is its minimum value.

For best accuracy of the predictive model, input variable selection is crucial. In this study, Partial Autocorrelation Function (PACF) was used to identify best inputs, utilizing two criteria for this purpose: selecting the input variable at lag t if its PACF value fell outside the nominal confidence interval of 95% and choosing previous value as an input if all PACF values were within a specified confidence interval.

Fig. 5 illustrates the PACF of the VMD-based sub-series of data for *DS5* (i.e., the training set for 2022 *EP* prediction). Based on this analysis, Eqs. (3) to (7) show the input variables for *DS1*, *DS2*, *DS3*, *DS4*, and *DS5*, respectively, where x_t = target (or output) variable and x_{t-p} = p antecedent variables of the target. It is worth mentioning that for all datasets (*DS1*, *DS2*, *DS3*, *DS4*, and *DS5*), the PACF of the RES sub-series fell within the confidence interval so this study has selected the previous electricity price value as an input variable for the designated models.

$$DS1 = \begin{cases} VMF1,\ 18\ (x_{t-1}, x_{t-2}) \\ VMF2-5,\ 16,\ 17\ (x_{t-1}, x_{t-2}, x_{t-3}) \\ VMF6-15\ (x_{t-2}, x_{t-4}) \\ RES(x_{t-1}) \end{cases} \tag{3}$$

$$DS2 = \begin{cases} VMF1,\ 17,\ 18\ (x_{t-1}, x_{t-2}) \\ VMF2-4,\ 16\ (x_{t-1}, x_{t-2}, x_{t-3}) \\ VMF5\ (x_{t-1}, x_{t-2}, x_{t-3}, x_{t-4}) \\ VMF6-15\ (x_{t-2}, x_{t-4}) \\ RES(x_{t-1}) \end{cases} \tag{4}$$

$$DS3 = \begin{cases} VMF1,\ 5,\ 15,\ 17\ (x_{t-1}, x_{t-2}) \\ VMF2-4,\ 16\ (x_{t-1}, x_{t-2}, x_{t-3}) \\ VMF6-14\ (x_{t-2}, x_{t-4}) \\ RES(x_{t-1}) \end{cases} \tag{5}$$

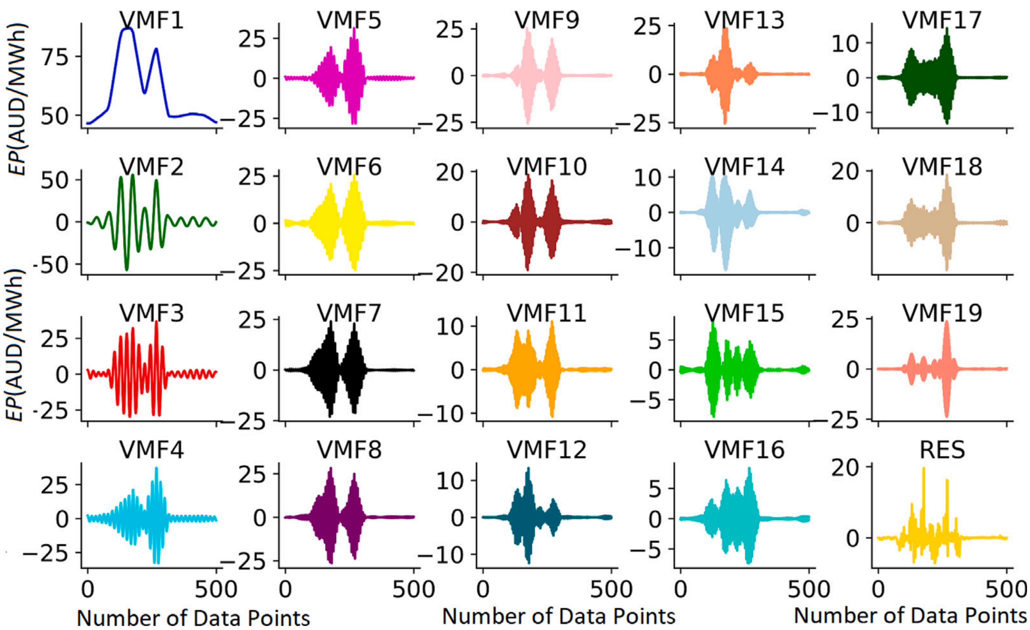


Fig. 4. The decomposition of *EP* using the VMD algorithm for *DS5* (training dataset for year 2022 *EP* prediction).

$$DS4 = \begin{cases} VMF1,\ 16,\ 17,\ 18,\ \&\ 19\ (x_{t-1}, x_{t-2}) \\ VMF2 - 5\ (x_{t-1}, x_{t-2}, x_{t-3}) \\ VMF6 - 15\ (x_{t-2}, x_{t-4}) \\ RES(x_{t-1}) \end{cases} \tag{6}$$

$$DS5 = \begin{cases} VMF1,\ 5,\ 15 \text{ and } 17\ (x_{t-1}, x_{t-2}) \\ VMF2,\ 3,\ 4 \text{ and } 16\ (x_{t-1}, x_{t-2}, x_{t-3}) \\ VMF6 - 13\ (x_{t-2}, x_{t-4}) \\ VMF14\ (x_{t-2}) \\ RES(x_{t-1}) \end{cases} \tag{7}$$

2.5.3. Preliminary prediction using VMD-CLSTM-VMD-ERCRF model

The VMD-CLSTM-VMD-ERCRF model was designed with a two-stage data decomposition system with an error compensation procedure (Fig. 6) with an overall predictive framework as follows:

The first stage was a preliminary prediction step using VMD-CLSTM-VMD-ERCRF, and the second stage was the error compensation technique. In respect to the first stage, the sub-series of VMD-based decomposition of *EP* series are denoted as VMF_1, VMF_2, VMF_3, and so on, up to VMF_n. Similarly, the second decomposition sub-series of residual errors (E_t) after the CLSTM model-based predictions are represented by $eVMF_1$, $eVMF_2$, $eVMF_3$, and so on, up to $eVMF_n$, with $eRES$ representing the residual component. The CLSTM model is utilized to make the initial prediction for each of the VMF-based data series.

In essence, CLSTM model takes in lagged normalized values of decomposed *EP* series as input with its output layer extracting the features for LSTM model. As input data consists of multivariate time series (a lagged matrix of decomposed *EP*, i.e. (X_{t-1}), (X_{t-2}), ……, (X_{t-n})), we define them as tensors with a shape of (N, Q, M) where N = number of samples, Q = maximum number of time steps across all variables and M = variables processed per time step. The numbers M_1 and M_2 denote the filters in the CNN layers, while Q_1 and Q_2 indicate the output dimensions of the LSTM layers. The output from LSTM layer goes through a flatten and a dense layer with a single neuron and linear activation function to predict the final outcome (X_t).

The accuracy of the proposed VMD-CLSTM-VMD-ERCRF model was improved by adopting the Bayesian optimization method [86] with the Gaussian process surrogate model to estimate the objective function based on prior experiments and an acquisition function to indicate the next input value to be evaluated. We determined parameter search range specified (Table B.12) to avoid over-fitting and under-fitting using "ReduceLROnPlateau" method while monitoring the loss based on MAE determined on a validation dataset. This followed the notion that when a loss variation of less than 5×10^{-3} in 5 consecutive epochs was noted, the learning rate was reduced by half until the minimum value of 10×10^{-6} was attained (see Fig. B.16.)

2.5.4. Error compensation technique

For the second stage, Fig. 5 shows the proposed error compensation technique purposely built to enhance the practicality of the proposed VMD-CLSTM-VMD-ERCRF model. The first step involves acquiring the error series of the training dataset from the original *EP* data subsequent to training the CLSTM network as per Eq. (8):

$$E(t) = EP_{CLSTM}(t) - EP_{act}(t), \tag{8}$$

where EP_{CLSTM} is the *EP* prediction during the training of the CLSTM network and EP_{act} is the actual *EP*.

Fig. 7 shows the training, validation and testing error series obtained by CLSTM for DS5. Similar to the previous method of predicting VMF with CLSTM, we used the VMD algorithm to decompose residual error series $E(t)$ and used a Random Forest (RF) model to predict each decomposed $eVMF$ component.

In respect to the choice of the algorithm for error compensation stage, it is noteworthy that the choice of an RF model for error predictions was made carefully by considering the strength of this method. In an RF model, the prediction speed is significantly faster than the training speed because we can save generated forests for future uses, which is actually a saving of computation cost on error calculations. The RF model is also able to handle the outliers (i.e. errors) by essentially binning them so it is indifferent to the non-linear features [87]. The RF model also has methods for balancing the errors in class population unbalanced data sets, as well as reduces over-fitting in decision trees to help to improve the accuracy. Furthermore, the averaging capability of an RF model makes it better than a single Decision Tree, or other models, and hence improves its accuracy and reduces over-fitting. In

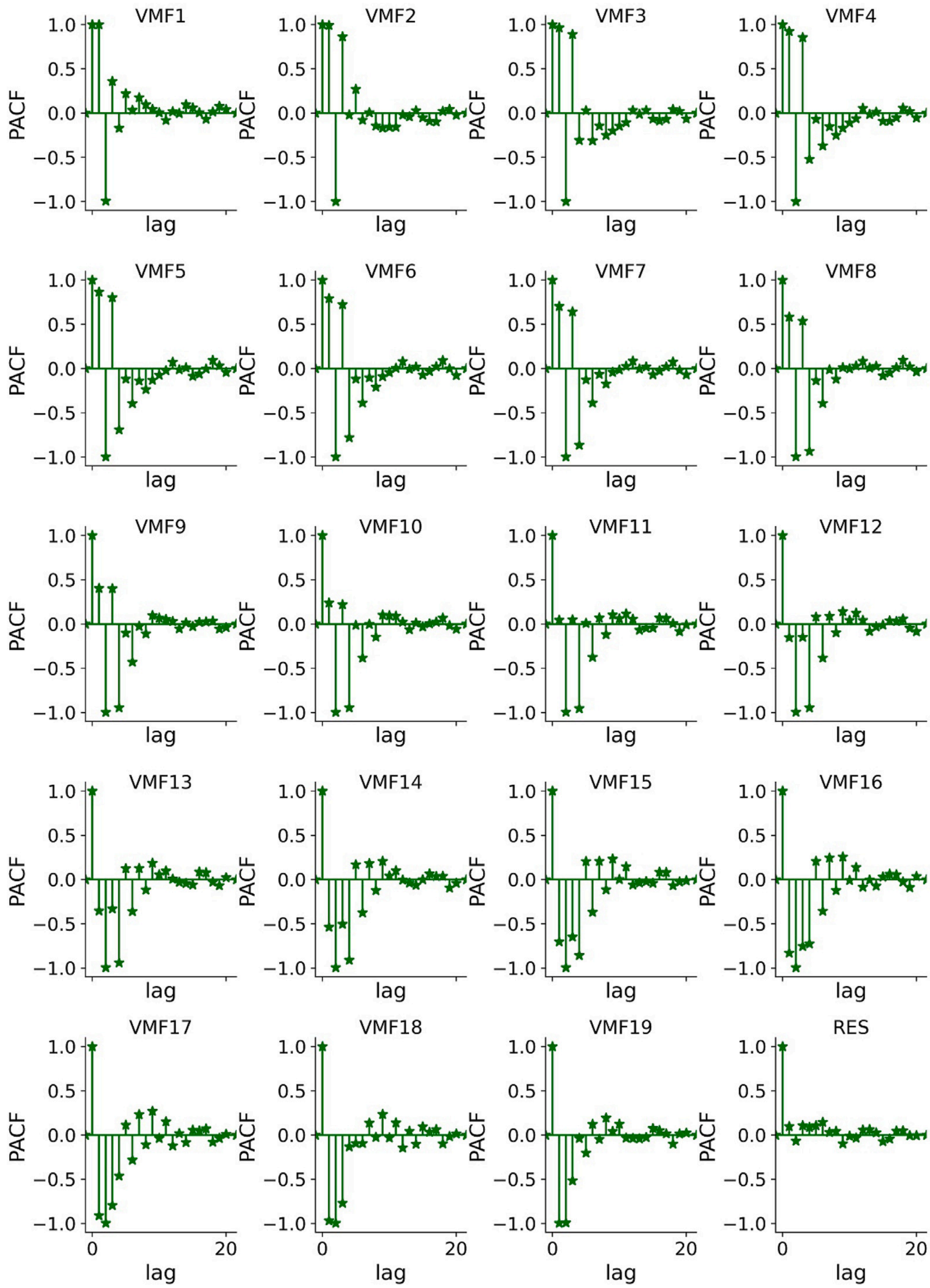


Fig. 5. PACF of VMFs and RES sub-series for*DS5* (training dataset for 2022 *EP* prediction).

our study, RF-based error predictions has led to significant performance improvement of the proposed VMD-CLSTM-VMD-ERCRF models.

The next task was to combine the prediction results of each $eVMF$ to obtain the final error prediction series. Once the RF model has predicted the error series, the final predicted half-hourly *EP* for the $DS1$, $DS2$, $DS3$, $DS4$ and $DS5$ can be obtained by Eq. (9):

$$EP_{fin}(t) = EP_{CLSTM}(t) + E_{RF}(t), \tag{9}$$

where $EP_{CLSTM}(t)$ is the CLSTM model prediction on testing dataset for half-hourly *EP* series and $E_{RF}(t)$ is the RF model's testing set prediction for the residual error series ($E(t)$) from Eq. (8).

To carry out VMD decomposition of residual error series, we set the parameters: α (quadratic penalty term), τ (noise tolerance), and ε (convergence criterion), to 2000, 0, and 10^{-7}, respectively. The number of sub-series (k) for VMD was determined using the CFR method ($K = 14, 15, 14, 17,$ *and* 14 for $DS1$, $DS2$, $DS3$, $DS4$ and $DS5$, *respectively*).

As an illustrated example we show for DS5, Fig. 8 displays the decomposed $eVMFs$ of residual error using VMD. The RF model's hyperparameters are fine-tuned using Bayesian optimization, with a primary focus on optimizing three parameters. These parameters include the number of trees in the forest (n _ estimators), the maximum depth of the tree (max _ depth), the number of features considered

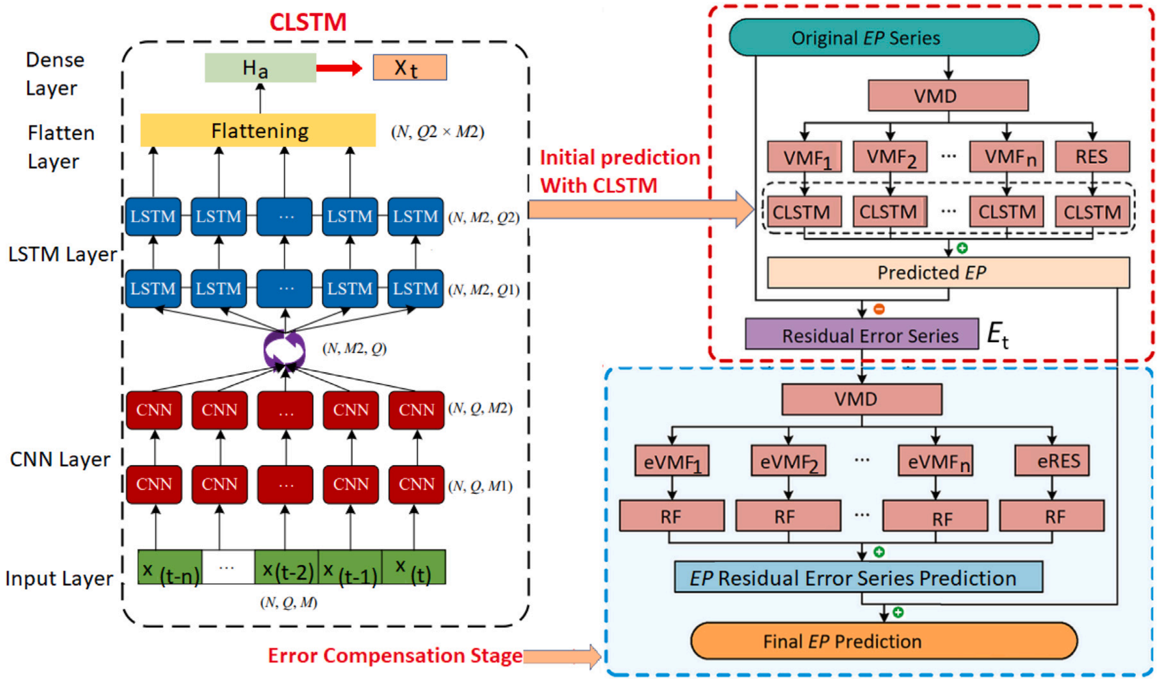


Fig. 6. The overall framework of the proposed VMD-CLSTM-VMD-ERCRF model.

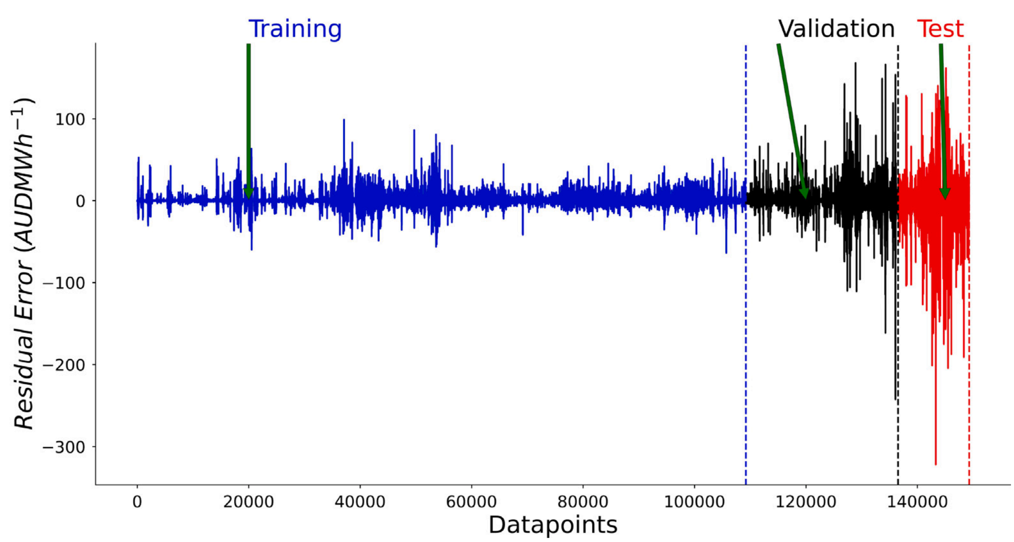


Fig. 7. Residual Error series obtained by CLSTM. (*Figure valid only for VMF1 training, Validation and testing of DS5*).

when searching for the optimal split (max _ features), and minimum number of samples to split an internal node (min _ samples _ split).

2.6. Benchmark models and performance evaluation criteria

To fully evaluate the proposed VMD-CLSTM-VMD-ERCRF model, five different predictive models (i.e., XGB, LSTM, RF, DNN, VMD-CLSTM) were adopted whose hyperparameters were optimized using Bayesian methods. For the applicable search range for hyperparameters, see Table B.12.

We adopted Mean Absolute Error (MAE):

$$MAE(\text{AUD/MWh}) = \frac{1}{N}\sum_{i=1}^{N} |EP^p - EP^a|, \quad (10)$$

Root Mean Square Error ($RMSE$):

$$RMSE(\text{AUD/MWh}) = \sqrt{\frac{1}{N}\sum_{i=1}^{N} (EP^p - EP^a)^2}, \quad (11)$$

Symmetric Mean Absolute Percentage Error ($sMAPE$):

$$sMAPE = \frac{1}{N}\sum_{i=1}^{N} \frac{|EP^a - EP^p|}{(|EP^a| + |EP^p|)/2}, \quad (12)$$

Legates and McCabe Index (I_{LM}):

$$I_{LM} = 1 - \frac{\sum_{i=1}^{N} |EP^p - EP^a|}{\sum_{i=1}^{N} |EP^a - \langle EP^a \rangle|}, \quad (13)$$

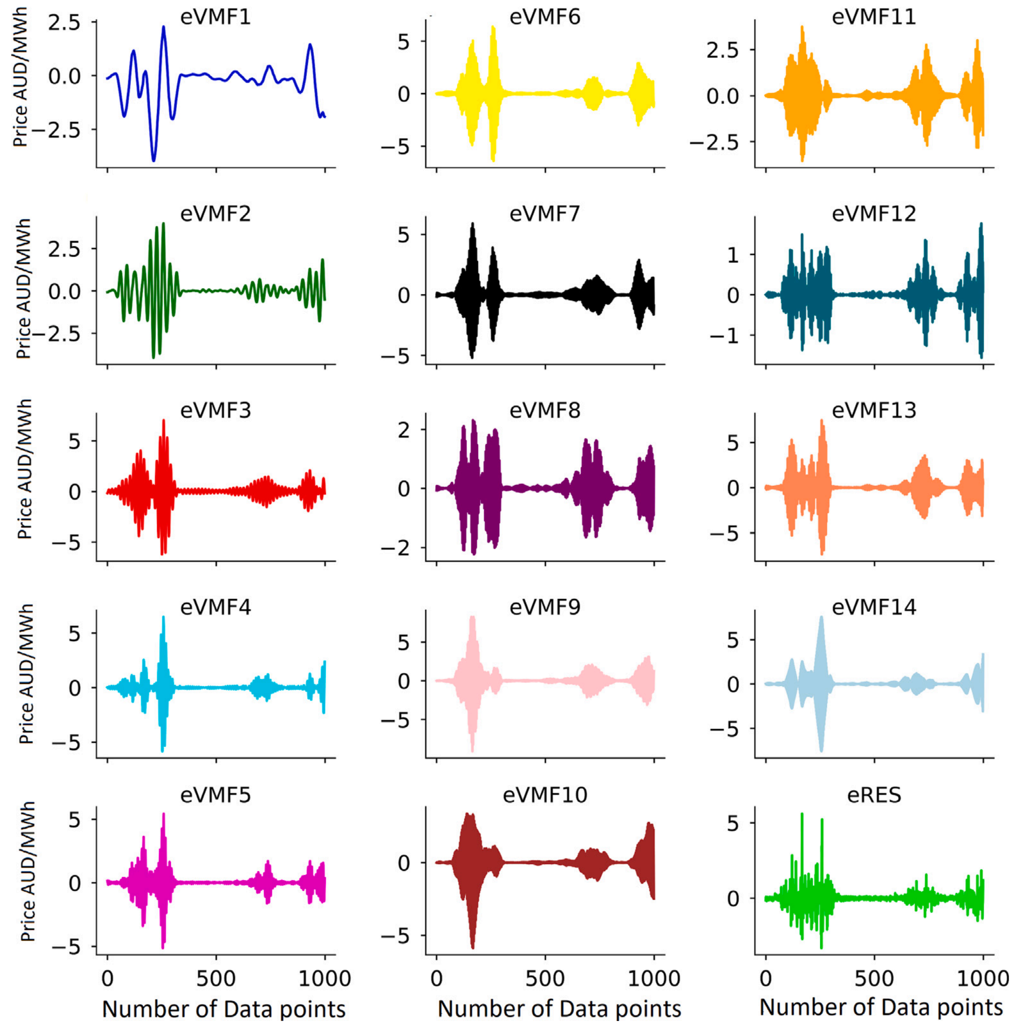


Fig. 8. CLSTM Residual Error VMF (eVMF) obtained by VMD.(for *DS5*).

Nash–Sutcliffe Index (I_{NS}):

$$I_{NS} = 1 - \frac{\sum_{i=1}^{N}(EP^a - EP^p)^2}{\sum_{i=1}^{N}(EP^a - \langle EP^a\rangle)^2}, \tag{14}$$

Wilmott's Index (I_{WI}):

$$I_{WI} = 1 - \frac{\sum_{i=1}^{N}(EP^a - EP^p)^2}{\sum_{i=1}^{N}(|EP^p - \langle EP^a\rangle| + |EP^o - \langle EP^p\rangle|)^2}, \tag{15}$$

Absolute Percentage Bias (APB):

$$\mathrm{APB}(\%) = \left|\frac{\sum_{i=1}^{n}(EP^a - EP^p)}{\sum_{i=1}^{n} EP^a}\right| \cdot 100, \tag{16}$$

Coefficient of Determination (R^2):

$$R^2 = \left(\frac{\sum_{i=1}^{N}(EP^a - \langle EP^a\rangle)(EP^p - \langle EPp\rangle)}{\sqrt{\sum_{i=1}^{n}(EP^a - \langle EP^o\rangle)^2}\sqrt{\sum_{i=1}^{n}(EP^p - \langle EPp\rangle)^2}}\right)^2, \tag{17}$$

and Skill Score (SS) for model evaluations:

$$SS = 1 - \frac{RMSE(p)}{RMSE(r)}. \tag{18}$$

Notable, EP^a and EP^p represent actual and predicted half-hourly *EP* while $\langle EP^a\rangle$ and $\langle EP^p\rangle$ represent actual and predicted mean *EP*, N = number of tested data points, $RMSE(p)$ and $RMSE(r)$ is the *RMSE* of prediction and reference models, respectively.

- The range of R^2 is [0, 1] while the MAE and $RMSE$ are measured in absolute units of *EP* (AUD/MWh) between [0, +∞] with 0 = perfect model and +∞ = poorly performing model. R^2 assesses the covariance to find out how well the modelled data fits actual data, whereas MAE and $RMSE$ measure the predictive power in absolute error terms.
- The range of I_{WI} is [0, 1], which is an improvement over $RMSE$ and MAE metrics. I_{WI} identifies both additive and proportional differences between observed and simulated means and variances.
- The range of I_{NS} is [−∞, 1] to assess the relative magnitude of residual variance compared to measured variance with a score of −∞ = worst fit and 1 indicating a perfectly fitted model.
- The range of I_{LM} is [0, 1]. This is a more robust metric compared to I_{NS} and I_{WI}, designed to overcome their limitations.
- The model with the lowest Symmetric Mean Absolute Percentage Error ($sMAPE$) is considered the best. $sMAPE$ is a symmetrical measure that avoids the problem of division by zero. In contrast, the conventional Mean Absolute Percentage Error($MAPE$) metric tends to become overinflated when the true value is close to zero, but $sMAPE$ does not have this issue.
- The Absolute Percentage Bias (APB) expresses the error of predicted values as a percentage relative to the observed values. A lower APB value, closer to zero, indicates good accuracy of the model. The optimal value for APB is zero.
- If the Skill Score (SS) is negative, the prediction is not better than the reference model. Conversely, if SS is positive, the prediction

Table 3
Evaluating the proposed VMD-CLSTM-VMD-ERCRF model for half-hourly EP predictions. R^2 = Coefficient of Determination, $RMSE$: AUD/MWh = Root Mean Square Error; MAE:AUD/MWh = Mean Absolute Error.

Dataset	Predictive models	R^2	$RMSE$	MAE
DS1 (Winter)	VMD-CLSTM-VMD-ERCRF	0.999	5.191	3.057
	VMD-CLSTM	0.874	11.352	6.770
	LSTM	0.882	45.171	22.781
	DNN	0.858	48.574	25.486
	XGB	0.883	44.846	22.902
	RF	0.882	44.983	23.050
DS2 (Autumn)	VMD-CLSTM-VMD-ERCRF	0.998	13.190	8.250
	VMD-CLSTM	0.916	19.412	13.124
	LSTM	0.911	59.306	34.654
	DNN	0.929	60.199	36.105
	XGB	0.897	61.038	36.375
	RF	0.930	60.123	35.595
DS3 (Spring)	VMD-CLSTM-VMD-ERCRF	0.998	6.247	3.388
	VMD-CLSTM	0.902	12.333	7.062
	LSTM	0.776	50.652	24.178
	DNN	0.779	50.510	24.704
	XGB	0.770	50.944	23.827
	RF	0.770	50.957	23.530
DS4 (Summer)	VMD-CLSTM-VMD-ERCRF	0.999	4.748	2.507
	VMD-CLSTM	0.891	11.367	5.900
	LSTM	0.832	48.753	20.855
	DNN	0.817	50.921	21.467
	XGB	0.839	47.923	20.223
	RF	0.837	48.168	19.497
DS5 (Yearly 2022 predictions)	VMD-CLSTM-VMD-ERCRF	0.999	10.220	5.154
	VMD-CLSTM	0.914	21.855	13.178
	LSTM	0.906	61.125	34.886
	DNN	0.872	63.370	37.680
	XGB	0.926	61.584	35.690
	RF	0.896	61.651	35.692

is an improvement over reference model. The degree of improvement is directly proportional to SS, meaning that higher scores imply greater enhancement.

Model selection using on single statistical metric is challenging, as each model has its own merits and constraints. We therefore adopted Global Performance Index (GPI) combining multiple metrics for a comprehensive evaluation where a higher GPI = greater accuracy. Rather than analysing individual metrics, the GPI, over a scaling of [0, 1], assigns equal weights to several statistical metrics [88]:

$$GPI_i = \sum_{j=1}^{n} \xi_j \left(\tilde{I}_i - I_{ij} \right), \tag{19}$$

where $\xi_j = -1$ for Pearson's Correlation Coefficient and $\xi_j = 1$ for all the other indicators. $\tilde{I}_i$ is the median of scaled values of indicator j and I_{ij} is the scaled value of indicator for model.

The interpretation of GPI is such that if a statistical metric is below median, a larger difference between that value and the median value of all other models indicates that the model is more accurate than other models. Similarly, if a statistical indicator's value is above the median, a greater deviation from the median indicates that the model is less accurate than other models.

This study also uses the promoting percentages (λ) of I_{WI} ($\lambda_{I_{WI}}$), I_{NS} ($\lambda_{I_{NS}}$), I_{LM} ($\lambda_{I_{LM}}$), and R^2 (λ_{R^2}), defined as follows:

$$\lambda_{I_{WI}} = \frac{|I_{WI(1)} - I_{WI(2)}|}{I_{WI(1)}}, \tag{20}$$

$$\lambda_{I_{NS}} = \frac{|I_{NS(1)} - I_{NS(2)}|}{I_{NS(1)}}, \tag{21}$$

$$\lambda_{I_{LM}} = \frac{|I_{LM(1)} - I_{LM(2)}|}{I_{LM(1)}}, \tag{22}$$

$$\lambda_{R^2} = \frac{|I_{R^2(1)} - I_{R^2(2)}|}{I_{R^2(1)}}. \tag{23}$$

The statistical corroboration between the actual and predicted half-hourly EP in the model's testing phase was investigated using the Diebold–Mariano (DM) statistic test. The DM statistic is defined as follows:

$$S_{DM} = \frac{\bar{g}}{\sqrt{\left(\hat{V}_g / N\right)}}, \tag{24}$$

where

$$\bar{g} = \left(\sum_{t=1}^{N} g_t \right) / N, \; g_t = \left(x_t - \hat{x}_{te,t} \right)^2 - \left(x_t - \hat{x}_{re,t} \right)^2, \tag{25}$$

and

$$\hat{V}_g = \gamma_0 + 2 \sum_{t=1}^{\infty} \gamma_t, \quad \left(\gamma_t = \mathrm{cov}\left(g_{t+1}, g_t \right) \right), \tag{26}$$

where γ_0 is the variance of g_t, $\hat{x}_{te,t}$ and $\hat{x}_{re,t}$ represent the predicted values of x_t calculated using the tested method te and reference method re, respectively, in period t. N is the number of observations in testing dataset.

Note that DM statistic aims to assess the significance of different models [89], to ascertain whether the expected forecast accuracy is equal across various models. This task uses $RMSE$ as a loss function with a null hypothesis that the $RMSE$ of the tested model (te) is not less than that of the reference model (re).

3. Result and discussion

This section presents the results of the proposed VMD-CLSTM-VMD-ERCRF and five benchmark models, for half-hourly electricity price predictions for Queensland, Australia evaluated for DS1 (winter), DS2 (Autumn), DS3 (Spring), DS4 (Summer), and DS5 (Yearly, 2022 prediction) periods, which refer to the Winter, Autumn, Spring and Summer seasons, and the Year 2022 dataset, respectively.

Table 3 demonstrates that the proposed VMD-CLSTM-VMD-ERCRF models outperforms all the other models with the highest R^2, the

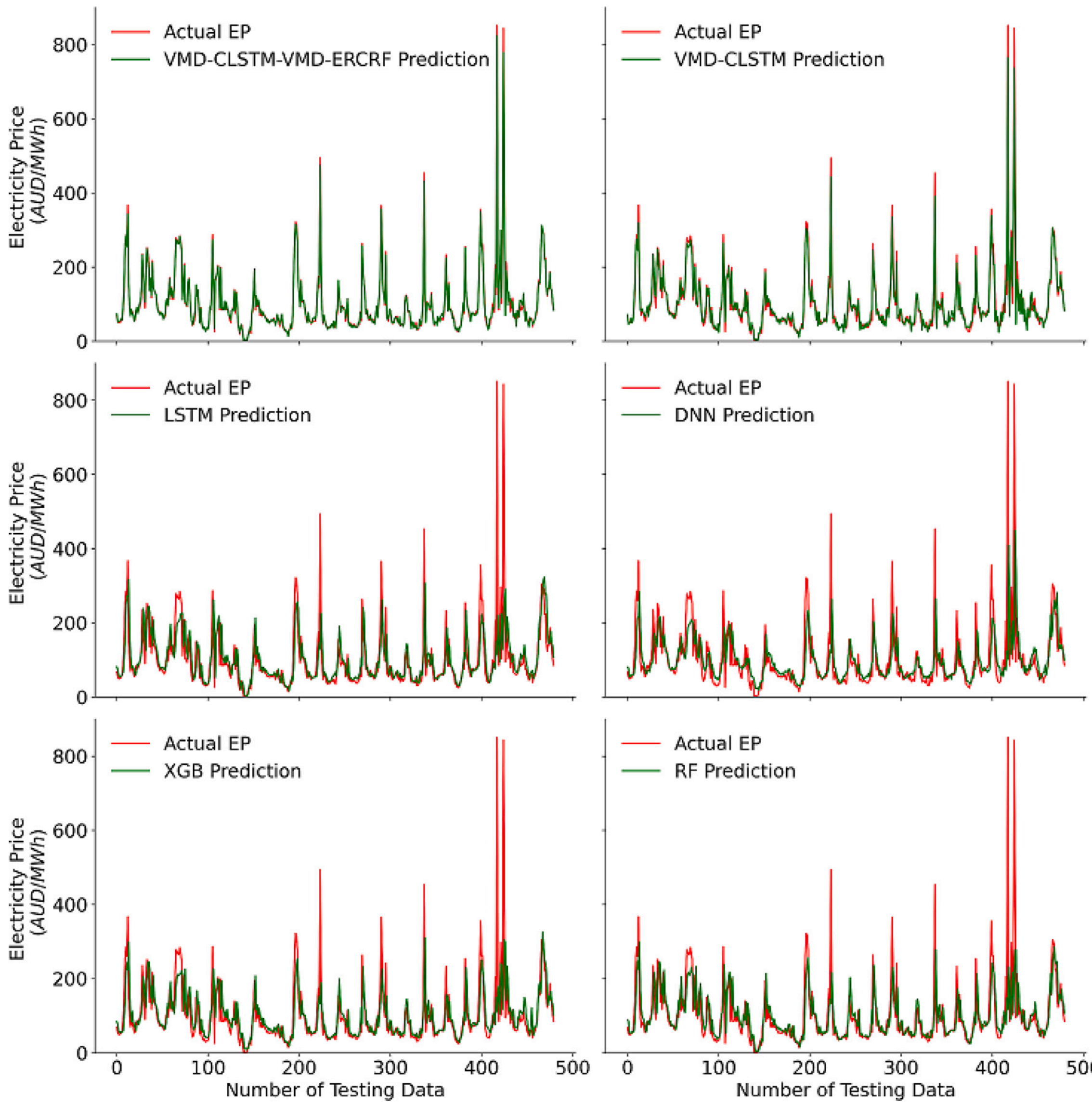


Fig. 9. Actual *vs.* model generated *EP* for the case of *DS1* (for conciseness, only 500 tested data points are shown.).

lowest MAE (2.5 to 8.25 AUD/MWh) and $RMSE$ (5.191 to 13.19 AUD/MWh) for all dataset. Note that here, the objective model is benchmarked against the VMD-CLSTM, XGB, RF, DNN and the LSTM models for four different seasons and the yearly (i.e., 2022) prediction dataset with the best model indicated in blue. For three datasets ($DS2$, $DS4$, and $DS5$), the R^2 measure indicated similar performance between the proposed model and benchmark models. However, the $RMSE$ and MAE values were lower for the proposed model at these sites when compared to the other benchmark models.

As an example, when predicting for the Autumn season ($DS2$), the R^2 values obtained were $\approx$ 0.998, $\approx$ 0.916, $\approx$ 0.911, $\approx$ 0.929, $\approx$ 0.897, and $\approx$ 0.930 for VMD-CLSTM-VMD-ERCRF, VMD-CLSTM, LSTM, DNN, XGB, and RF models, respectively. In contrast, the $RMSE$ values were $\approx$ 13.190, $\approx$ 19.412, $\approx$ 59.603, $\approx$ 60.199, $\approx$ 61.038, and $\approx$ 60.123, and the MAE values were $\approx$ 8.250, $\approx$ 13.124, $\approx$ 34.654, $\approx$ 36.105, $\approx$ 36.375, and $\approx$ 35.595 for VMD-CLSTM-VMD-ERCRF, VMD-CLSTM, LSTM, DNN, XGB, and RF models, respectively.

In addition, the models that utilized both CLSTM and data decomposition (VMD-CLSTM-VMD-ERCRF and VMD-CLSTM) demonstrated superior prediction performance compared to single models (LSTM, DNN, XGB, and RF). This indicates that using data decomposition is a useful method to improve the accuracy of predictions. However, it is important to acknowledge that the R^2 metric is not affected by scale and offset, which could result in higher values for sub-optimal models. Additionally, the $RMSE$ measure may have a bias towards high predicted values due to the squaring of residuals. In contrast, the MAE measure does not provide information on whether the model overestimates or underestimates since it only considers the absolute value, thus making these three measures potentially unreliable when comparing models with similar underlying structures. Therefore, to overcome the limitations of R^2, $RMSE$, and MAE in accurately evaluating models with similar structures, normalized error measures such as I_{WI}, I_{NS}, and I_{LM} were used in this study.

The Nash–Sutcliffe Index (I_{NS}) is a widely used evaluation metric that is a scaled version of MSE and is dimensionless. However, it tends to exaggerate the impact of larger outliers while ignoring smaller ones. To address this issue, Willmott's Index (I_{WI}) was introduced, which considers the ratio of MSE instead of differences. The results of the I_{NS} and I_{WI} metrics for predicting *EP* at a half-hourly interval for five datasets ($DS1$, $DS2$, $DS3$, $DS4$, and $DS5$) are presented in Table 4.

It is clear that the DL model integrated with the 2-stage decomposition with Error Compensation (VMD-CLSTM-VMD-ERCRF) exhibited a significant enhancement, as both I_{NS} and I_{WI} values were higher than those of the standalone models. As an example, when predicting for the

Table 4
Evaluating the proposed VMD-CLSTM-VMD-ERCRF model for half-hourly *EP* predictions using the normalized, non-dimensional model evaluation metrics: I_{WI} = Wilmott's Index; I_{NS} = Nash–Sutcliffe Index; I_{LM} = Legates and McCabe Index for four different seasons and the yearly (i.e., 2022) prediction dataset with the best model indicated in blue.

Dataset	Predictive models	I_{WI}	I_{NS}	I_{LM}
DS1 (Winter)	**VMD-CLSTM-VMD-ECRF**	**0.994**	**0.995**	**0.935**
	VMD_CLSTM	0.860	0.845	0.785
	LSTM	0.644	0.602	0.514
	DNN	0.548	0.539	0.456
	XGB	0.625	0.607	0.511
	RF	0.638	0.605	0.508
DS2 (Autumn)	**VMD-CLSTM-VMD-ECRF**	**0.992**	**0.988**	**0.916**
	VMD_CLSTM	0.882	0.873	0.767
	LSTM	0.839	0.752	0.648
	DNN	0.825	0.744	0.634
	XGB	0.817	0.738	0.631
	RF	0.826	0.746	0.639
DS3 (Spring)	**VMD-CLSTM-VMD-ECRF**	**0.983**	**0.990**	**0.897**
	VMD_CLSTM	0.832	0.862	0.716
	LSTM	0.377	0.358	0.266
	DNN	0.298	0.362	0.250
	XGB	0.369	0.351	0.277
	RF	0.364	0.350	0.286
DS4 (Summer)	**VMD-CLSTM-VMD-ECRF**	**0.992**	**0.995**	**0.930**
	VMD_CLSTM	0.877	0.882	0.835
	LSTM	0.513	0.477	0.415
	DNN	0.489	0.429	0.398
	XGB	0.512	0.494	0.433
	RF	0.506	0.489	0.453
DS5 (Yearly 2022 predictions)	**VMD-CLSTM-VMD-ECRF**	**0.996**	**0.994**	**0.953**
	VMD_CLSTM	0.858	0.877	0.819
	LSTM	0.851	0.801	0.680
	DNN	0.828	0.786	0.654
	XGB	0.845	0.798	0.672
	RF	0.844	0.798	0.672

Table 5
Evaluating the proposed VMD-CLSTM-VMD-ERCRF model for half-hourly *EP* predictions using Absolute Percentage Bias APB: % and Symmetric Mean Absolute Percentage Error $sMAPE$: % for four different seasons and the yearly (i.e., 2022) prediction dataset with the best model indicated in blue.

Dataset	Predictive models	*APB*	*sMAPE*
DS1 (Winter)	**VMD-CLSTM-VMD-ERCRF**	**3.66%**	**5.83%**
	VMD-CLSTM	8.11%	11.14%
	LSTM	27.31%	28.05%
	DNN	30.55%	31.61%
	XGB	27.45%	28.03%
	RF	27.63%	27.96%
DS2 (Autumn)	**VMD-CLSTM-VMD-ERCRF**	**4.27%**	**4.08%**
	VMD-CLSTM	6.79%	7.40%
	LSTM	17.93%	19.94%
	DNN	18.68%	20.52%
	XGB	18.82%	21.37%
	RF	18.41%	20.36%
DS3 (Spring)	**VMD-CLSTM-VMD-ERCRF**	**4.55%**	**6.40%**
	VMD-CLSTM	9.49%	11.81%
	LSTM	32.48%	30.79%
	DNN	33.18%	31.09%
	XGB	32.00%	29.34%
	RF	31.61%	28.86%
DS4 (Summer)	**VMD-CLSTM-VMD-ERCRF**	**2.55%**	**2.79%**
	VMD-CLSTM	6.00%	6.42%
	LSTM	21.19%	19.21%
	DNN	21.81%	19.28%
	XGB	20.55%	18.14%
	RF	19.81%	17.52%
DS5 (Yearly 2022 predictions)	**VMD-CLSTM-VMD-ERCRF**	**2.61%**	**3.08%**
	VMD-CLSTM	6.67%	8.44%
	LSTM	17.66%	20.14%
	DNN	19.07%	21.61%
	XGB	18.06%	20.22%
	RF	18.06%	20.04%

Table 6
Evaluating the proposed VMD-CLSTM-VMD-ERCRF model for half-hourly *EP* predictions using Diebold–Mariano (DM) test statistic with the best model indicated in blue. DS1 (Winter); DS2 (Autumn); DS3 (Spring); DS4 (Summer); DS5 (Yearly 2022 predictions).

Models	*DS1*	*DS2*	*DS3*	*DS4*	*DS5*
VMD-CLSTM-VMD-ERCRF	**0.991**	**0.964**	**0.988**	**0.993**	**0.978**
VMD-CLSTM	0.955	0.921	0.955	0.961	0.898
LSTM	0.290	0.262	0.234	0.281	0.202
DNN	0.296	0.242	0.224	0.298	0.188
XGB	0.300	0.219	0.225	0.306	0.190
RF	0.179	0.240	0.238	0.216	0.142

Winter season ($DS1$), the I_{WI} values obtained were ≈ 0.994, ≈ 0.860, ≈ 0.644, ≈ 0.548, ≈ 0.625, and ≈ 0.638 for VMD-CLSTM-VMD-ERCRF, VMD-CLSTM, LSTM, DNN, XGB, and RF models, respectively. Similarly, the I_{NS} values were ≈ 0.995, ≈ 0.845, ≈ 0.602, ≈ 0.539, ≈ 0.607, and ≈ 0.605 for VMD-CLSTM-VMD-ERCRF, VMD-CLSTM, LSTM, DNN, XGB, and RF models, respectively.

The I_{NS} and I_{WI} results provide evidence that the DL model integrated with the 2-stage decomposition with Error Compensation (VMD-CLSTM-VMD-ERCRF) model resulted in enhanced performance of the standalone deep learning model (DNN and LSTM) for all Dataset. It is important to note that the metrics I_{NS} and I_{WI} may overemphasize peak residual values and yield inflated scores due to the squared residuals.

Conversely, Legates and McCabe Index (I_{LM}) is not biased towards overestimating errors and discrepancies as it utilizes absolute values and appropriate weights. Hence, I_{LM} is regarded as a more dependable measure than I_{NS} and I_{WI}. The model evaluation based on I_{LM} consistently demonstrates that the VMD-CLSTM-VMD-ERCRF model outperformed the other models across all five datasets. The proposed model achieved I_{LM} scores greater than 0.897 for all datasets, with the highest score of 0.953 recorded for $DS5$ (Year 2022 prediction) dataset. Since I_{NS}, I_{WI} and I_{LM} values are greater than 0.90 for six datasets, the VMD-CLSTM-VMD-ERCRF model can be considered as a well performed model to estimate half-hourly *EP*.

To evaluate model bias, percentage error measures were utilized in the study, namely, Absolute Percentage Bias (APB) and Symmetric Mean Absolute Percentage Error ($sMAPE$). Although APB is a popular and easily understandable measure, it has limitations, such as being highly influenced by a few outliers and having no upper limit. To address these limitations, $sMAPE$ was developed, which uses the average of the prediction and observed of comparison as the denominator, providing an upper limit of 200% and a well-defined range to assess relative errors. Hence, $sMAPE$ is considered to have several theoretical advantages over APB, whose denominator is based solely on the standard of comparison.

The proposed VMD-CLSTM-VMD-ERCRF model demonstrated the lowest APB and $sMAPE$ values for all datasets in predicting half-hourly *EP*, as shown in Table 5. For example, during the summer season ($DS4$), the proposed model (VMD-CLSTM-VMD-ERCRF) generated an APB of ≈ 2.55% and $sMAPE$ of ≈ 2.79%, which were lower than the values produced by VMD-CLSTM (≈ 6.67% and ≈ 8.44%), LSTM (≈ 17.66% and ≈ 20.14%), DNN (≈ 19.07% and ≈ 21.61%), XGB (≈ 18.06% and ≈ 20.22%), and RF (≈ 18.06% and ≈ 20.04%).

Table 5 demonstrate that the VMD-CLSTM-VMD-ERCRF model produces significantly improved prediction for the $DS1$, $DS2$, $DS3$, and $DS5$ compared to other models such as VMD-CLSTM, LSTM, DNN, XGB, and RF. These findings indicate that VMD-CLSTM-VMD-ERCRF can be a reliable and effective tool for half-hourly *EP* prediction. Moreover, since the variation of *EP* over short time periods (i.e., half-hourly) is relatively consistent, the persistence model is usually used as a baseline model for calculating the Skill Score (SS). This model assumes that the

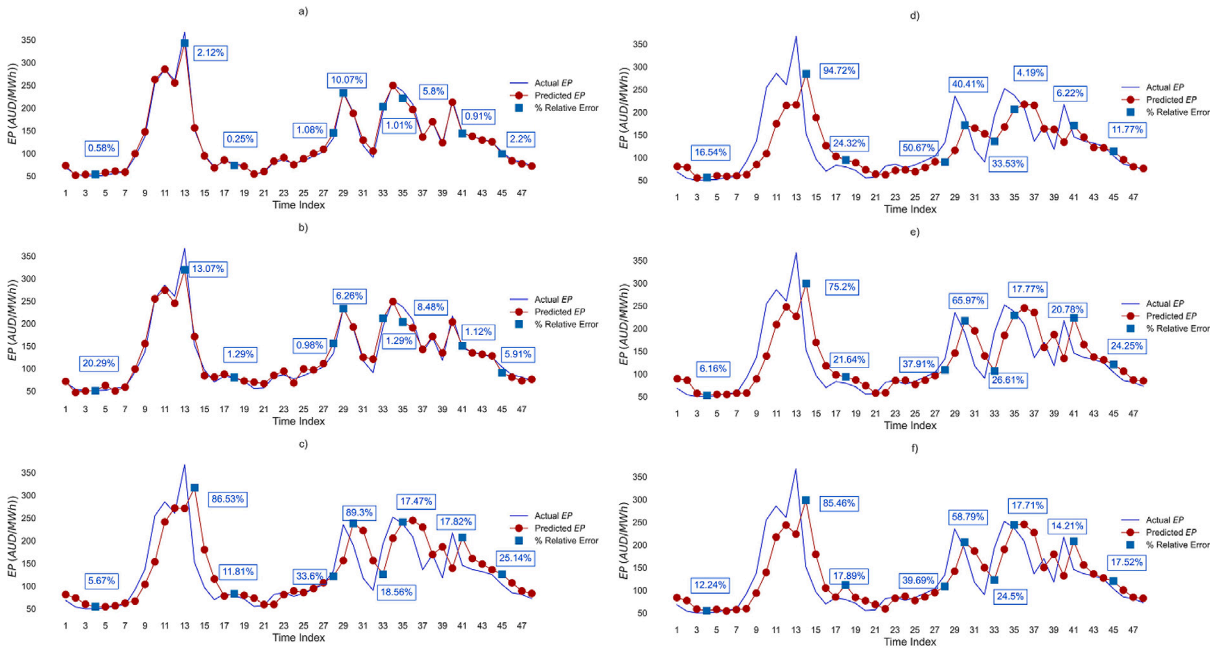


Fig. 10. (a) Actual *vs.* predicted half-hourly *EP* generated by the half-hourly VMD-CLSTM-VMD-ERCRF hybrid model in the testing phase, shown for 1-day dataset for the case of $DS1$. Comparison models are: (b)VMD-CLSTM, (c) LSTM, (d) DNN, (e) XGB and (f) RF. The relative error encountered is indicated in the blue colour.

predicted value $EP(t+T)$ at time T ahead is equal to the current value $EP(t)$.

Table 6 shows the SS values for different prediction models, and as expected, all of the models perform better than the persistence model. The VMD-CLSTM-VMD-ERCRF model achieved the best results with an SS value of ≈ 0.991, ≈ 0.964, ≈ 0.988, ≈ 0.993, and ≈ 0.978 for $DS1$, $DS2$, $DS3$, $DS4$, and $DS5$, respectively.

Comparing the SS for other models, we observe significant improvements induced by the decomposition and error compensation mechanism, ranging from an increase in skill score value from $SS \approx 0.955$ to 0.991 ($DS1$), $SS \approx 0.921$ to 0.964 ($DS2$), $SS \approx 0.955$ to 0.988 ($DS3$), $SS \approx 0.961$ to 0.993 ($DS4$), and $SS \approx 0.898$ to 0.978 ($DS5$) when comparing the VMD-CLSTM model to the proposed VMD-CLSTM-VMD-ERCRF model. Additionally, the standalone models (LSTM, DNN, XGB, and RF) exhibit SS values in the range of 0.142 to 0.290. In summary, the decomposition-based model can significantly improve performance, especially when combined with error compensation.

To illustrate visually the degree of similarity between predicted and actual electricity price, we used a line plot as shown in Fig. 9 as well as Figs. C.18 to C.21 in Appendix C, to compare the prediction results for half-hourly *EP* generated by the proposed VMD-CLSTM-VMD-ERCRF versus VMD-CLSTM, LSTM,DNN, XGB and RF benchmark models. The graph reveals that the predictions produced by VMD-CLSTM-VMD-ERCRF model are more similar to the actual data compared with the other models. This similarity is particularly evident for extreme values, such as the 12nd, 151st, 223rd, 337th, 417th, 424th, 466th point of $DS1$, 32nd, 84th, 132nd 270th, 278th, 367th, 411st, 463rd point of $DS2$, 22nd, 70th, 160th, 283rd, 421st, 463rd, 496th point of $DS3$, 36th, 83rd, 252nd, 345th, 393rd, 437th, 489th point of $DS4$ and 37th, 83rd, 123rd, 224th, 271st, 321st, 363rd, 368th, 439th, 449th, 464th point of $DS5$ case. These indicate that the VMD-CLSTM-VMD-ERCRF model performs well even for high and fluctuating electricity prices data.

Not surprisingly, while the benchmark models (VMD-CLSTM, LSTM, DNN, XGB, RF) perform well in terms of predictability, they still struggle to maintain accurate predictions for high *EP* values, which is a disadvantage of these relatively inferior models. This is exemplified in Fig. 10. In contrast, the proposed VMD-CLSTM-VMD-ERCRF model demonstrates significantly superior performance in predicting the peak electricity price data when compared with VMD-CLSTM, LSTM, DNN, XGB and RF models. Specifically, the VMD-CLSTM-VMD-ERCRF model only underestimates peak values by 2.12%, whereas the VMD-CLSTM, LSTM, DNN, XGB and RF models underestimate them by 13.07%, 86.53%, 94.72%, 75.2%, and 85.46% respectively for the $DS1$. These results further demonstrate the suitability of the VMD-CLSTM-VMD-ERCRF model for half-hourly *EP* prediction.

Table 7 shows the promoting percentages (λ) based on the Coefficient of Determination, Willmott's Index, Nash–Sutcliffe and Legates and McCabe Index computed from the predicted and actual electricity price datasets in the testing phase. Here, the objective model is benchmarked against the VMD-CLSTM, XGB, RF, DNN and the LSTM models for four different seasons and the yearly (i.e., 2022) prediction dataset. The table demonstrates the efficacy of the proposed VMD-CLSTM-VMD-ERCRF model over benchmark models, calculated in respect to R^2, I_{WI}, I_{NS}, and I_{LM} metrics for half-hourly *EP* prediction. In fact, the value of λ_{R^2}, $\lambda_{I_{WI}}$, $\lambda_{I_{NS}}$, and $\lambda_{I_{LM}}$ of the VMD-LSTM was 12.46%, 13.47%, 15.07%, and 15.97%, respectively, for the DS1 dataset, compared with the proposed VMD-CLSTM-VMD-ERCRF model. Similarly, the λ_{R^2}, $\lambda_{I_{WI}}$, $\lambda_{I_{NS}}$, and $\lambda_{I_{LM}}$ value of LSTM, DNN, XGB, and RF models relative to the proposed VMD-CLSTM-VMD-ERCRF model shows that the proposed model had a higher percentage of promotion, indicating that VMD-CLSTM-VMD-ERCRF outperformed the others in this *EP* prediction problem.

Table 8 shows the results of the Diebold–Mariano (DM) statistical test performed to compare the performance of the proposed VMD-CLSTM-VMD-ERCRF model with the other models for each of the datasets. Note that a positive DM value is expected to indicate significantly better performance of the VMD-CLSTM-VMD-ERCRF model compared to the other models in respect to *EP* predictions. Therefore, both the λ and the DM tests provide complementary evidence that the proposed VMD-CLSTM-VMD-ERCRF model outperformed all benchmark models.

Diagnostic plots were created to examine the model's prediction errors (PE) whereby ideally, a PE value should be zero for the best performing model and their distribution should be as close to zero as possible. To make the results easy to interpret, we present the absolute prediction error ($|PE|$) quantities. Fig. 11 illustrates the superior prediction capability of the proposed VMD-CLSTM-VMD-ERCRF model for

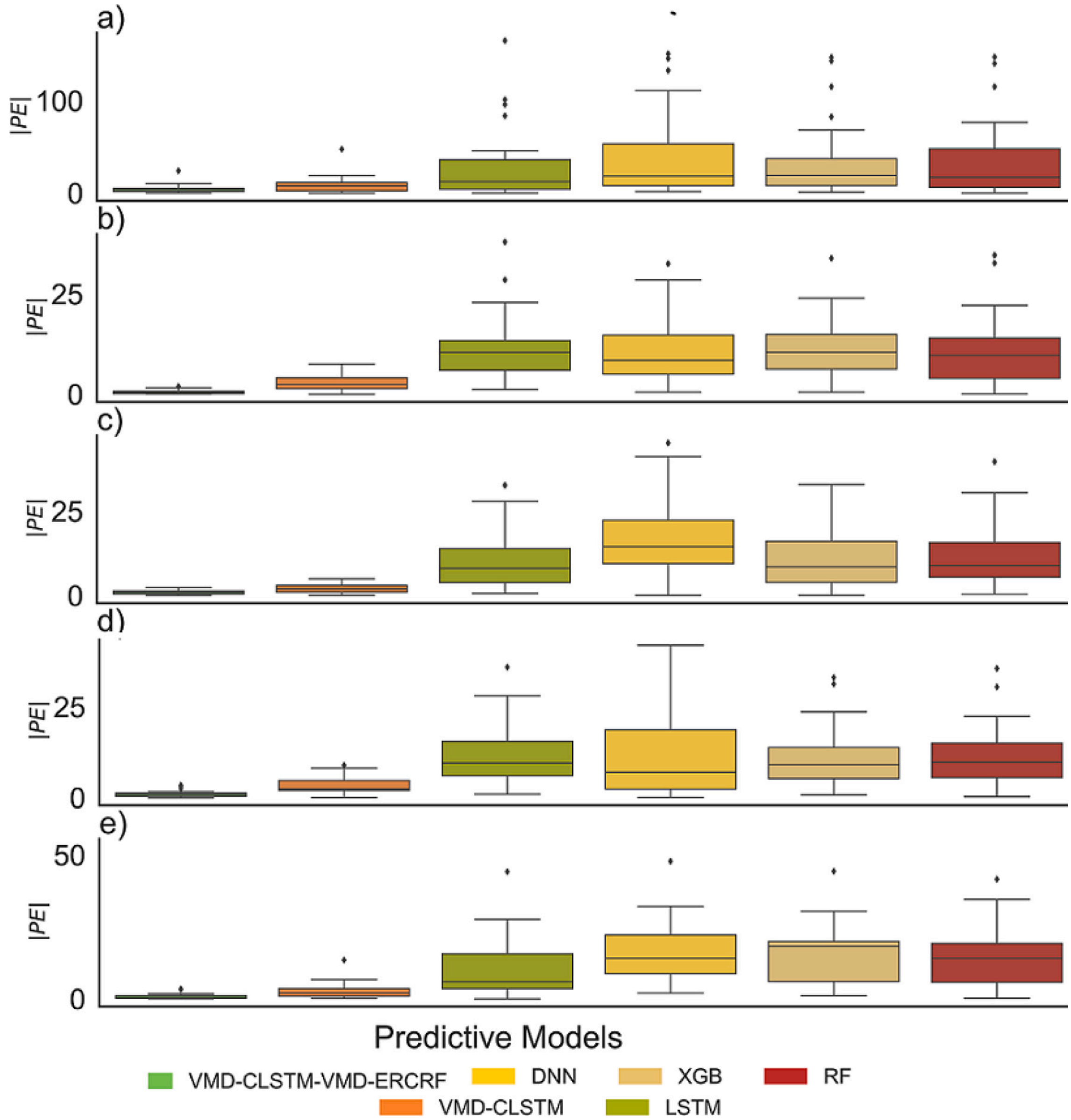


Fig. 11. Boxplots of errors computed between predicted and actual *EP* using the proposed VMD-CLSTM-VMD-ERCRF *vs.* VMD-CLSTM, LSTM, DNN, XGB and RF models in the testing phase for for four different seasons (a-d) and the yearly (i.e., 2022) prediction dataset (e).

all seasons as it exhibited smaller PE divisions compared to the other models. This concurs with the results in Tables 3–7.

Fig. 12 shows the Empirical Cumulative Distribution Function ($ECDF$), which provides a clear view of the distribution of $|PE|$. Importantly, the $ECDF$ for VMD-CLSTM, LSTM, DNN, XGB and RF showed very similar profiles but in contrast, that of VMD-CLSTM-VMD-ERCRF model was notably narrow, confined within a smaller range. Therefore, based on Fig. 10 and the $ECDF$ plots in Fig. 11, the proposed VMD-CLSTM-VMD-ERCRF model exhibited a superior performance in predicting half-hourly *EP*.

We now show the Taylor diagram (Fig. 13), which graphically depicts the relationship between the Standard Deviation (SD), Root Mean Square Deviation ($RMSD$), and Correlation Coefficient (r) of the predicted and actual electricity price data to showcase the strength of the proposed model. Accordingly, the proposed VMD-CLSTM-VMD-ERCRF model appears the closest to the observed value (i.e., OBS), indicating the best performance.

Although evaluation metrics and diagnostic plots were utilized to compare the models, ranking a large number of models based on such metrics, which have their own merits and constraints, can be challenging. To overcome this, a robust global performance indicator (GPI) was used. Fig. 14 displays the GPI, which shows that the proposed VMD-CLSTM-VMD-ERCRF models outperformed VMD-CLSTM, LSTM, DNN, XGB, and RF models in terms of performance ($GPI \approx 8.901$ ($DS1$), 9.317 ($DS2$), 4.229 ($DS3$), 13.154 ($DS4$), and 8.417 ($DS5$)). In general, the proposed VMD-CLSTM-VMD-ERCRF model had the highest GPI and the best predictive performance, ranking as the top model for all five datasets.

3.1. Comparison of model's computational complexity

This study includes a comparison of the computational complexity of the models as computation time is an essential factor to consider for practical application of any model. Table 9 shows the time taken for hyperparameter optimization, model training, and testing for half-hourly *EP* prediction for six different forecast models. The experiment was conducted using a Dell Precision 7920 with Intel Core i7-6700k CPU, and a parallel algorithm applied during the decomposition processes and training and testing of the models.

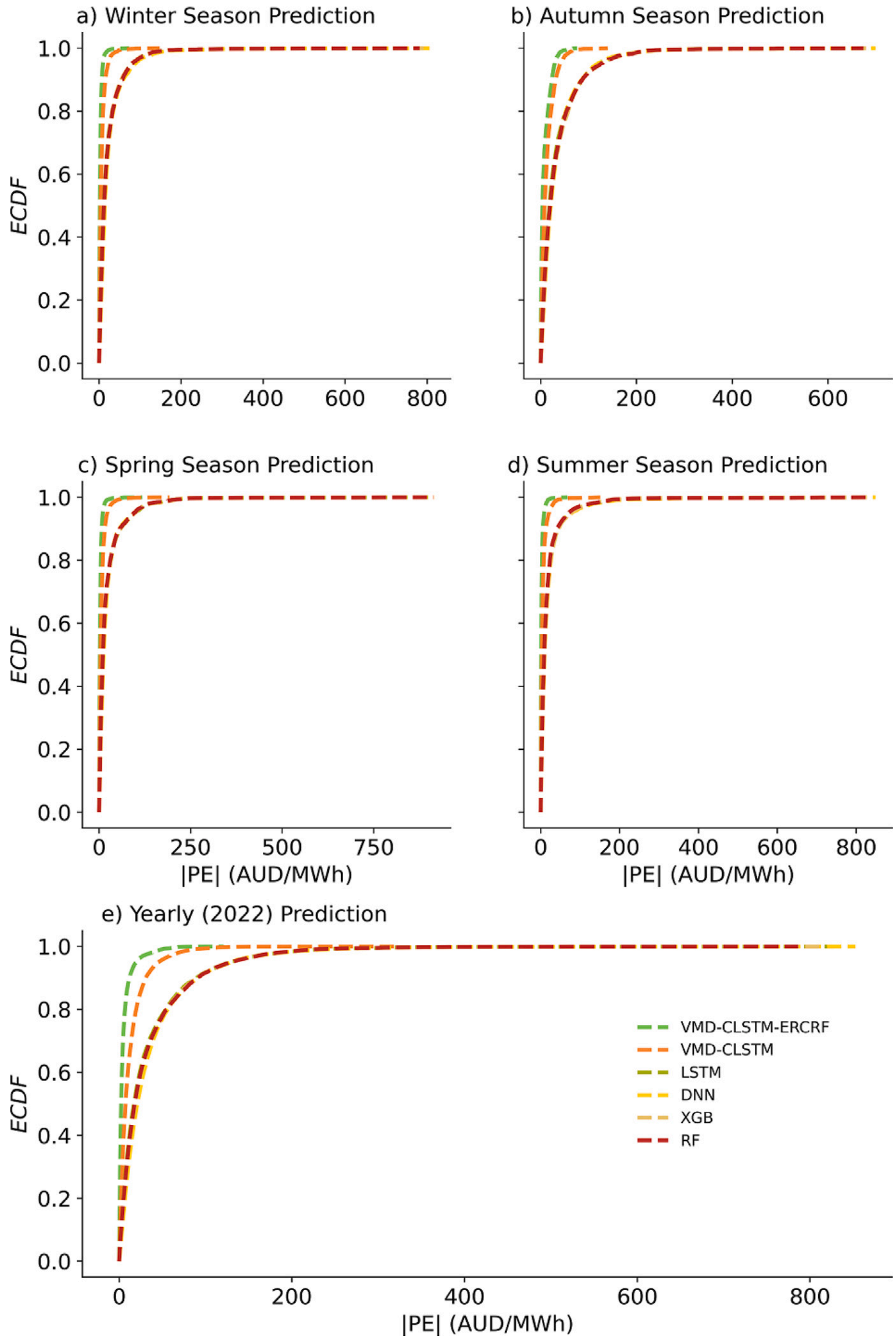


Fig. 12. Empirical cumulative distribution function ($ECDF$) of the Prediction Error (PE) generated by the proposed VMD-CLSTM-VMD-ERCRF *vs.* VMD-CLSTM, LSTM, DNN, XGB and RF models for four different seasons (a-d) and the yearly (i.e., 2022) prediction dataset (e).

The results show that the average time taken for hyperparameter optimization of the proposed VMD-CLSTM-VMD-ERCRF model was ≈ 6.18 h, which is relatively long. However, once the optimal hyperparameters were identified, this model can be used for an extended period. The testing data length varies for the five datasets, with $DS5$ having the longest period (i.e., the Year 2022 predictions), which affects the computation time. Nonetheless, the testing calculation time is less than 20 s, indicating that the proposed model is applicable to practical situations. Furthermore, the hyperparameter optimization and training time for *EP* prediction can be significantly reduced with advancements in hardware and software environments as well as code optimization. This reduction in training time is highly beneficial for practical applications.

4. Conclusions, limitations and recommendations for future research

4.1. Conclusions

Electricity price prediction is a critical part of electricity market. This research introduced a two-stage data decomposition and predictive modelling strategy combining time series prediction with an error compensation strategy. The proposed VMD-CLSTM-VMD-ERCRF hybrid model was verified for half-hourly electricity price predictions using real sub-station data for Queensland, Australia. The first phase of the model architecture was an initial prediction method while second phase was an error compensation stage. In the initial stage, the electricity price data were decomposed using Variational Mode Decomposition

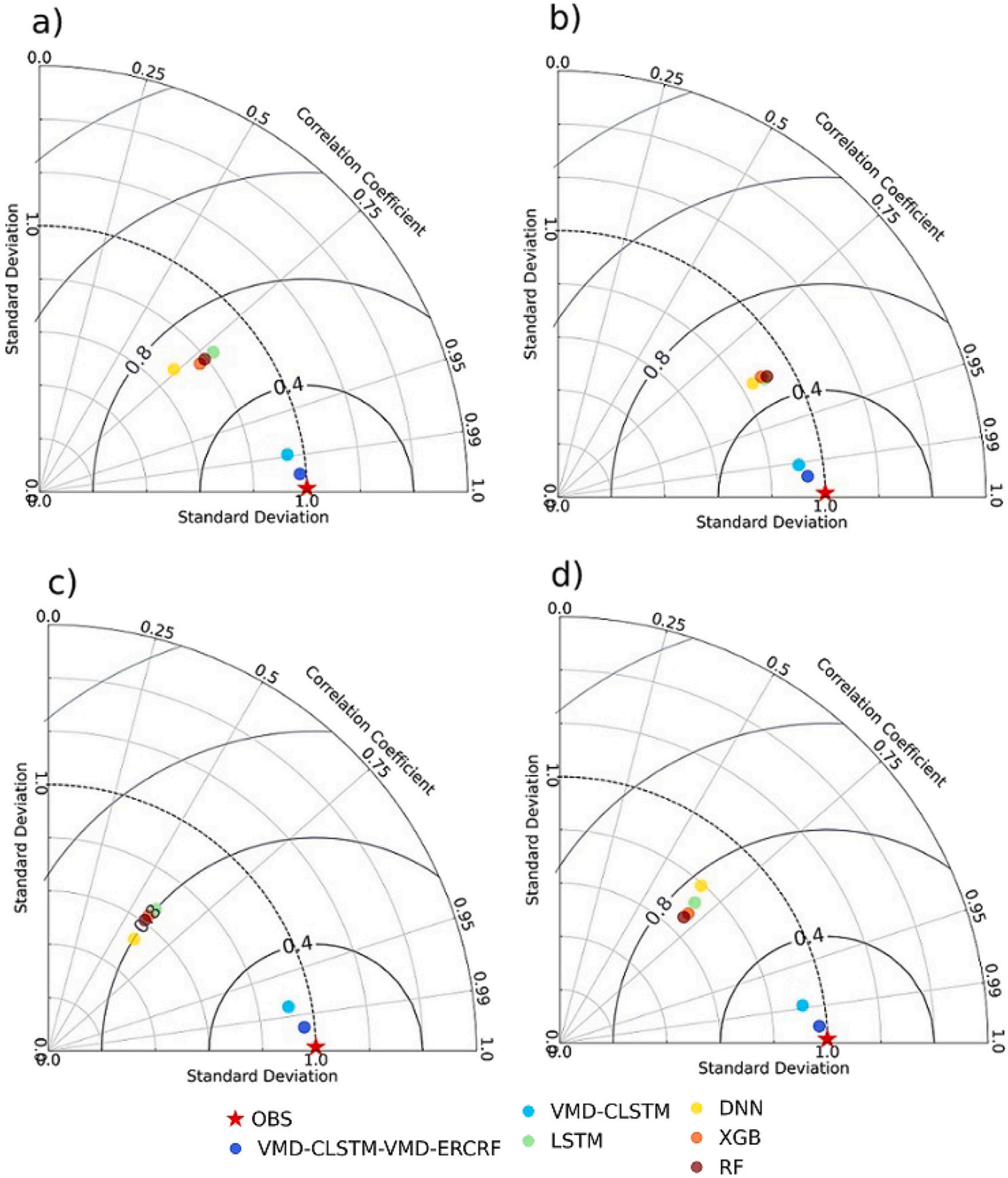


Fig. 13. Taylor diagram depicting correlation coefficients of the proposed VMD-CLSTM-VMD-ERCRF *vs.* VMD-CLSTM, LSTM, DNN, XGB and RF models for four different seasons (a-d) and the yearly (i.e., 2022) prediction dataset (e).

method, a sequential process to decompose input signal into a discrete number of sub-signals known as modes where each mode had a limited bandwidth but represented distinct patterns, features, periodicity trends and other stochastic or chaotic behaviours found in electricity price data. To the develop proposed VMD-CLSTM-VMD-ERCRF hybrid model, Partial Autocorrelation Functions were employed to extract the significantly lagged features of each of the intrinsic mode functions used later as an input for the proposed prediction model.

Firstly, the hybrid CNN-LSTM model was developed as a predictor framework for initial electricity price prediction. In the second phase, the error series of the initial predictions were collected with Variational Mode Decomposition method applied to further decompose the error series, leading to an enhancement in the overall capability of the proposed VMD-CLSTM-VMD-ERCRF model. An RF model was applied to this system to predict each of the VMD error series with the initial prediction results and error prediction results combined to finalize the electricity price prediction model. To fully ascertain the efficacy of the method, the proposed VMD-CLSTM-VMD-ERCRF hybrid model was verified over electricity price data split into five distinct sets: $DS1$ for Winter season prediction, $DS2$ for Autumn season prediction, $DS3$ for Spring season prediction, $DS4$ for Summer season prediction and $DS5$ for the year 2022 prediction, as shown in Table 2.

Five competing comparison models fully ascertained the efficacy of the proposed VMD-CLSTM-VMD-ERCRF hybrid model. One of these (VMD-CLSTM) used the data decomposition method without an error compensation stage while the other four (LSTM, DNN, XGB, RF) used data decomposition and error compensation altogether, providing a large pool of predictive models for a detailed evaluation of our objective model. The comprehensive analysis of results conclude that the proposed VMD-CLSTM-VMD-ERCRF model had superior performance,

Table 7
Evaluating the proposed VMD-CLSTM-VMD-ERCRF model for half-hourly *EP* predictions using Promoting percentages based on Coefficient of Determination (λ_{R^2}), Wilmott's Index ($\lambda_{I_{WI}}$), Nash–SutcliffeIndex ($\lambda_{I_{NS}}$) and Legates and McCabe Index ($\lambda_{I_{LM}}$).

Dataset	Predictive models	λ_{R^2}	$\lambda_{I_{WI}}$	$\lambda_{I_{NS}}$	$\lambda_{I_{LM}}$
DS1 (Winter)	VMD_CLSTM	12.46%	13.47%	15.07%	15.97%
	XGB	11.60%	37.09%	38.93%	45.34%
	RF	11.69%	35.83%	39.18%	45.68%
	LSTM	11.70%	35.18%	39.52%	45.06%
	DNN	14.08%	44.84%	45.78%	51.24%
DS2 (Autumn)	VMD_CLSTM	8.29%	11.07%	11.57%	16.31%
	LSTM	8.71%	15.43%	23.89%	29.23%
	RF	6.81%	16.72%	24.52%	30.27%
	DNN	6.95%	16.82%	24.65%	30.83%
	XGB	10.11%	17.67%	25.33%	31.13%
DS3 (Spring)	VMD_CLSTM	9.64%	15.30%	12.96%	20.23%
	DNN	21.94%	69.66%	63.46%	72.10%
	LSTM	22.26%	61.61%	63.83%	70.32%
	XGB	22.79%	62.42%	64.57%	69.13%
	RF	22.88%	62.93%	64.61%	68.13%
DS4 (Summer)	VMD_CLSTM	10.76%	11.65%	11.40%	10.24%
	XGB	16.04%	48.39%	50.31%	53.45%
	RF	16.21%	48.99%	50.84%	51.26%
	LSTM	16.76%	48.30%	52.09%	55.35%
	DNN	18.25%	50.75%	56.87%	57.20%
DS5 (Year 2022 predictions)	VMD_CLSTM	8.50%	13.83%	11.76%	14.03%
	LSTM	9.25%	14.53%	19.45%	28.64%
	XGB	7.29%	15.12%	19.73%	29.42%
	RF	10.28%	15.20%	19.76%	29.42%
	DNN	12.67%	16.86%	20.94%	31.33%

Table 8
Evaluation the VMD-CLSTM-VMD-ERCRF against benchmark models using Diebold–Mariano (DM) test statistic for half-hourly *EP* predictions. The column of the table is compared with rows and if the result is positive, the model in the rows is superior to the one in the column; otherwise, if it is negative, the one in the column is superior. Note: The top-performing model is indicated in bold (blue) and the objective model is benchmarked against the VMD-CLSTM, XGB, RF, DNN and the LSTM models for four different seasons and the yearly (i.e., 2022) prediction dataset.

Dataset	Predictive models	VMD-CLSTM	LSTM	DNN	XGB	RF
DS1 (Winter)	**VMD-CLSTM-VMD-ERCRF**	**5.60**	**5.49**	**5.58**	**5.41**	**5.53**
	VMD-CLSTM		5.47	5.57	5.38	5.51
	LSTM			5.14	−1.58	−0.80
	DNN				−5.72	−5.14
	XGB					0.71
DS2 (Autumn)	**VMD-CLSTM-VMD-ERCRF**	**11.53**	**9.88**	**9.47**	**10.38**	**10.23**
	VMD-CLSTM		9.68	9.27	10.20	10.05
	LSTM			1.83	6.27	2.90
	DNN				1.59	−0.16
	XGB					−3.67
DS3 (Spring)	**VMD-CLSTM-VMD-ERCRF**	**5.13**	**3.58**	**3.84**	**3.50**	**3.28**
	VMD-CLSTM		3.52	3.78	3.44	3.22
	LSTM			−0.17	0.91	0.39
	DNN				0.43	0.30
	XGB					0.02
DS4 (Summer)	**VMD-CLSTM-VMD-ERCRF**	**5.07**	**4.59**	**4.59**	**4.78**	**4.43**
	VMD-CLSTM		4.56	4.56	4.75	4.40
	LSTM			2.88	−1.36	−0.76
	DNN				−2.91	−2.09
	XGB					0.22
DS5 (Year 2022 predictions)	**VMD-CLSTM-VMD-ERCRF**	**12.57**	**14.77**	**14.01**	**14.82**	**14.84**
	VMD-CLSTM		14.64	13.84	14.69	14.72
	LSTM			4.78	2.15	2.33
	DNN				−3.83	−3.78
	XGB					0.37

achieving the highest R^2 values and the lowest MAE from 2.5–8.25 AUD/MWh and $RMSE$ from 5.191–13.19 AUD/MWh across all tested datasets, as per Table 3, made a substantial improvement compared to benchmark models, evidenced by higher I_{NS} and I_{WI} as per Table 4, and showed the best performance in terms of the lowest APB and $sMAPE$, which is evidenced in Table 5.

The proposed VMD-CLSTM-VMD-ERCRF model also outperformed all benchmark models for half-hourly *EP* prediction, as indicated by higher λ values. For instance, for the $DS1$ dataset, VMD-CLSTM had λ values of 12.46%, 13.47%, 15.07%, and 15.97% for R^2, I_{WI}, I_{NS}, and I_{LM}, respectively, compared to VMD-CLSTM-VMD-ERCRF. See Table 7.

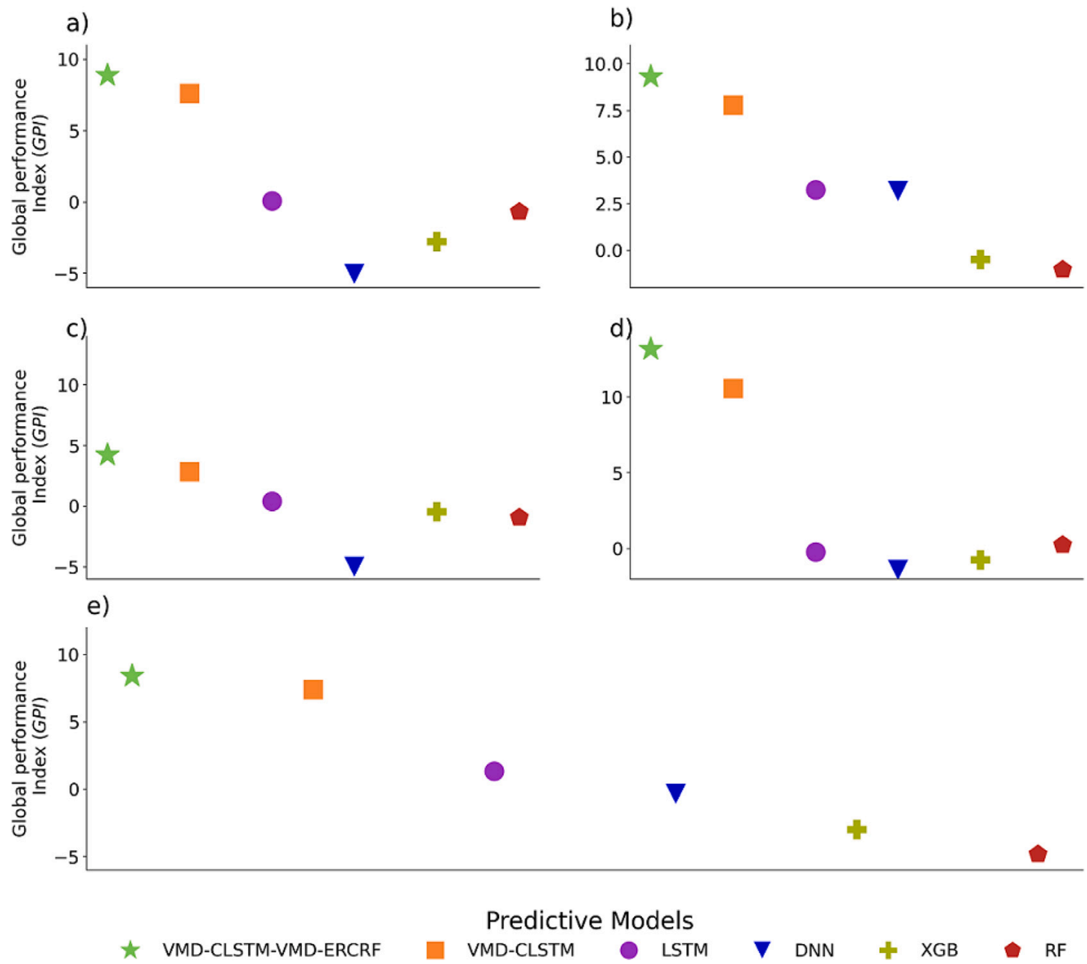


Fig. 14. Model ranking using the Global Performance Index (GPI). (a) $DS1$, (b) $DS2$, (c) $DS3$, (d) $DS4$, and (e) $DS5$.

Table 9
Average computation time of all models.

Dataset	Designated model	Hyperparameter optimization (h)	Model construction time (Model training validation) (min)	Prediction time (Testing) (s)
DS1, DS2, DS3, and DS4	VMD-CLSTM-VMD-ERCRF	6.180	37	18
	VMD-CLSTM	5.230	33	14
	LSTM	5.100	18	12
	DNN	3.210	17	8
	XGB	1.210	11	4
	RF	1.520	11	4
DS5	VMD-CLSTM-VMD-ERCRF	8.520	49	20
	VMD-CLSTM	7.530	38	21
	LSTM	6.240	21	15
	DNN	4.250	21	10
	XGB	1.560	14	6
	RF	1.325	15	6

Regarding statistical testing, the DM test indicated that the performance margin between the proposed VMD-CLSTM-VMD-ERCRF and five benchmark models (VMD-CLSTM, LSTM, DNN, XGB, RF) was statistically significant to further ascertain the robustness of the metrics used to arrive at a superior performance outcome. See Table 8.

In accordance with the above findings, the insights based on visual comparisons of predicted and actual electricity price and related model metrics show that the proposed VMD-CLSTM-VMD-ERCRF model outperforms all state-of-the-art deep learning models. As a result, electricity price data are captured in terms of high- and low-frequency perturbations which are non-stationary and nonlinear. Compared with previous studies that used wavelet transform (WT), EMD, or variants of it (EEMD, CEEMD), the proposed method was less susceptible to noise. As a result of hyperparameter optimization and feature selection, the final model was fine-tuned to achieve lower error metrics and improved prediction performance by compensating for sampling errors as part of the current VMD method. An important insight was the importance of assessing model errors more closely. In order to accomplish this, the original electricity price series was decomposed into subseries and a residual series was calculated with the VMD algorithm and LSTM network. By comparing the predicted subseries with the original observation value, an error series was constructed.

Using VMD to decompose the error series into sub-series, which were then predicted using a Random Forest network to compensate for the prediction results of the original series, the half-hourly electricity price was highly accurate. To predict the future value of electricity prices, the VMD-CLSTM-VMD-ERCRF model considered the relatively complex antecedent electricity price time series. Following these insights and contributions, we conclude that VMD-CLSTM-VMD-ERCRF, based on error compensation strategies achieved through model input data decomposition, can improve prediction accuracy.

4.2. Limitations, opportunities and recommendations for future research work

In spite of the outstanding performance, the wider interpretability of the VMD-CLSTM-VMD-ERCRF model for different electricity markets remains an important aspect and open problem of investigation that requires further attention. The model was tested with datasets aggregated together for a single State in Australia (i.e. Queensland) without considering how it would behave in a geographically diverse region. Thus, model testing should be conducted in different geographical regions with wider changes in electricity prices. By using more diverse datasets, the VMD-CLSTM-VMD-ERCRF model can be tested for different behaviours of electricity prices. As well as social, geopolitical, and economic factors, a revised new modelling scheme should be tested in terms of its interpretability for diverse input datasets.

The present VMD-CLSTM-VMD-ERCRF model was tested on a single variable (i.e., investigating antecedent electricity price) to build a forecast system. In future, one could develop a fully explainable (xAI) and interpretable model by using methods like Local Interpretable Model-Agnostic Explanations (LIME), Shapley additive explanations (SHAP), and permutation feature importance (PFI) or a Bayesian optimized ensemble Neural Basis Expansion Analysis for Interpretable Time Series (B-E-NBEATS) method [90,91]. In addition to revealing relationships between model inputs and predictions, these methods provide a greater physical understanding of how the VMD-CLSTM-VMD-ERCRF model could arrive at a particular prediction. Thus, the causes and effects of electricity prices could be explained better through the consideration of climatic, social, geopolitical, and economic factors. In addition, data decomposition and error correction strategies can improve the predictive performance and physical understanding of these input variables, their fluctuations, as well as the impact on better versions of the VMD-CLSTM-VMD-ERCRF model.

For this study, we have built a VMD-CLSTM-VMD-ERCRF model using half-hourly electricity prices. However, Australia's National Electricity Market (NEM) operated by the Australian Energy Market Operator (AEMO) calculates the electricity price based on the 5–30 arrangement where five-minute dispatch prices are averaged to produce a 30-min Trading Price or "spot price" that is then used to settle purchase and sale transactions. As of October 1st, 2021, the NEM switched from a 30-min settlement to a 5-min settlement so that users could better adjust their consumption in response to electricity price changes [92, 93]. This change made it easier to control appliances during a 5-min, high price period instead of a 30-min high price period for businesses and households. In addition to this, the 30-min settlement usage also demonstrates limitations in metering and data handling technologies, so a 5-min period would encourage lower electricity costs. A future study could investigate how to build the VMD-CLSTM-VMD-ERCRF model using the 5-min settlement datasets, and how to integrate the data with real-time weather and weather events that affect electricity demand, as well as electricity market stability, such as bush fires and storms.

Through the improved VMD-CLSTM-VMD-ERCRF model, with its five minute prediction capability, we could be able to invest efficiently in new technologies such as batteries, which can be used to back up wind and solar power, and consumers can participate in the market more efficiently by responding to demand and generators responding to demand at a much granular level (in real-time). By using an error correction method, the improved VMD-CLSTM-VMD-ERCRF model can align the market's price signal with the physical electricity system The improved price signals forecasts can lead to more efficient decisions by generators, lowering wholesale costs and typical electricity bills over time. Finally, the model's wider applicability could be tested by expanding its half-hourly forecast horizon. In order to test the effectiveness of the VMD-CLSTM-VMD-ERCRF model over these timescales, future studies should use hourly, daily, weekly, monthly as well as long-term (yearly) datasets.

CRediT authorship contribution statement

Sujan Ghimire: Writing – original draft, Software, Investigation, Data curation, Conceptualization. **Ravinesh C. Deo:** Writing – review & editing, Validation, Supervision, Resources, Project administration, Investigation, Funding acquisition, Conceptualization. **David Casillas-Pérez:** Writing – review & editing, Visualization, Investigation, Conceptualization. **Sancho Salcedo-Sanz:** Writing – review & editing, Supervision, Investigation, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data were acquired from AEMO. (https://www.aemo.com.au/).

Acknowledgements

The authors thank data providers, all reviewers and Editor for their thoughtful comments, suggestions and review process. Partial support of this work was through a project PID2020-115454GB-C21 of the Spanish Ministry of Science and Innovation (MICINN).

Appendix A. List of acronyms

Tables A.10 and A.11 show the abbreviations and acronyms used in his paper.

Appendix B. Theoretical overviews

B.1. Theory of VMD method

Let $\{EP(t)\}_{t=1}^{H}$ denote a typically non-stationary electricity price sequence of *EP* discrete values sampled at periodic intervals. Then, its decomposition using VMD can be expressed by Eq. (B.1):

$$EP(t) = \sum_{k=1}^{K} F_k(t) + r(t) \tag{B.1}$$

where $F_k(t)$ is the kth IMF sequence, K is the decomposition level, and $r(t)$ is the residual. Adapting [84], IMFs are amplitude-modulated and frequency-modulated signals following Eq. (B.2):

$$F_k(t) = A_k(t)\cos\phi_k(t), \quad A_k(t) \geq 0 \tag{B.2}$$

where $\phi_k(t)$ is defined as phase, $A_k(t)$ is the envelope corresponding to the kth IMF. It also has a slowly varying instantaneous frequency that is mostly compact around a central frequency ω_k.

Through the use of optimization techniques, the VMD algorithm finds the K IMFs and their respective central frequencies concurrently. The constrained variational optimization problem, expressed

Table A.10
List of acronyms.

Acronym	Expansion
AI	Artificial Intelligence
ANN	Artificial Neural Network
AR	Autoregressive
ARIFMA	Autoregressive Fractionally Integrated Moving Average
ARIMA	Autoregressive Integrated Moving Average
ARMA	Autoregressive Moving Average
AUD	Australian Dollar
BDL	Bayesian Deep Learning
BiLSTM	Bi-Directional LSTM
BPNN	Back-Propagation Neural Network
Catboost	Categorical Boosting
CEEMD	Complementary Ensemble Empirical Mode Decomposition
CEEMDAN	Complementary Ensemble Empirical Mode Decomposition with Adaptive Noises
CNN	Convolutional Neural Network
DBN	Deep Belief Network
DELM	Deep Extreme Learning Machine
DL	Deep Learning
ELM	Extreme Learning Machine
EMD	Empirical Mode Decomposition
ENN	Elman Neural Networks
EP	Electricity Prices
ERC-DNN	Error Compensation Deep Neural Network
ES	Exponential Smoothing
EWT	Empirical Wavelet Transform
GARCH	Generalized Autoregressive Conditional Heteroskedasticity
GPR	Gaussian Process Regression
GRNN	Generalized Regression Neural Network
GRU	Gated Recurrent Unit
ICEEMDAN	Improved Complementary Ensemble Empirical Mode Decomposition with Adaptive Noises
IDPSO	Inverted and Discrete Particle Swarm Optimization
ILRCNN	Integrated Long-Term Recurrent Convolutional Network
KDE	Kernel Density Estimate
KNN	K-Nearest Neighbours
KNNR	K-Nearest Neighbours Regression
LASSO	Least Absolute Shrinkage and Selection Operator

Table A.11
List of acronyms.

Acronym	Expansion
ADMM	Alternating Direction Method of Multipliers
LR	Linear Regression
LSTM	Long Short-Term Memory
MARS	Multivariate Adaptive Regression Splines
MIF	Mutual Information
ML	Machine Learning
MLP	Multilayer Perceptron
NARMAX	Nonlinear Autoregressive Moving Average Model With Exogenous Inputs
NBEATS	Neural Basis Expansion Analysis For Interpretable Time Series
PACF	Partial Autocorrelation Function
PDF	Probability Density Function
PI	Prediction Interval
PNN	Probabilistic Neural Network
RBFNN	Radial Basis Function Neural Network
RF	Random Forest Regression
RM	Regression Model
RNN	Recurrent Neural Networks
SEQ	South-East Queensland
SHAP	Shapley Additive Explanations
SILO	Scientific Information For Land Owners
SM	Statistical Methods
SRM	Structural Risk Minimization
SSA	Sparrow Search Algorithm
STL	Seasonal and Trend Decomposition Using Loess
SVR	Support Vector Regression
TCN	Temporal Convolutional Network
TF	Transfer Function
VMD	Variational Mode Decomposition
WNN	Wavelet Neural Network
XGB	eXtreme Gradient Boos

by Eq. (B.3), can be solved using alternating direction method of multipliers (ADMM).

$$\min_{\{u_k\},\{\omega_k\}} \left\{ \sum_k \|\partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) \otimes u_k(t) \right] e^{-j\omega_k t} \|_2^2 \right\} \quad s.t. \quad \sum_k u_k(t) = f(t) \tag{B.3}$$

where $\delta(t)$ denotes the Dirac distribution, $\otimes$ and ∂ denotes convolution and partial differential operators, respectively, and $\{u_k\} := \{u_1, \ldots, u_k\}$ and $\{\omega_k\} := \{\omega_1, \ldots, \omega_k\}$ represents the IMF and central frequencies, respectively. Eq. (B.3) can be addressed by introducing a quadratic penalty and Lagrangian multipliers ($\lambda(t)$). The augmented Lagrangian (ℓ) is given by :

$$\ell\left(\{u_k\}, \{\omega_k\}, \lambda\right) := \alpha \sum_k \|\partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \|_2^2 + \|f(t) - \sum_k u_k(t)\|_2^2 + \left\langle \lambda(t), f(t) - \sum_k u_k(t) \right\rangle \tag{B.4}$$

where α denotes the equilibrium parameter of the data-fidelity constraint, and the term:

$$\left\| f(t) - \sum_k u_k(t) \right\|_2^2, \tag{B.5}$$

is the quadratic penalty term to accelerate the convergence rate. The modes $u_k(\omega)$ in the frequency domain are estimated using ADMM in the form of the Wiener filter structure as Eq. (B.6):

$$\hat{u}_k^{n+1}(\omega) = \frac{\hat{f}(\omega) - \sum_{i \neq k} \hat{u}_i + \frac{\hat{\lambda}(w)}{2}}{1 + 2\alpha\left(\omega - \omega_k\right)^2}, \tag{B.6}$$

where $\hat{u}_k$, $\hat{f}(\omega)$, $\hat{\lambda}(w)$, and $\hat{u}_i$ are the Fourier transform (FT) of the components with n the number of iterations. The central frequency (ω_k) are updated using Eq. (B.7) and the λ is simultaneously updated by Eq. (B.8).

$$\hat{\omega}_k^{n+1} = \frac{\int_0^\infty \omega \left|\hat{u}_k(\omega)\right|^2 d\omega}{\int_0^\infty \left|\hat{u}_k(\omega)\right|^2 d\omega} \tag{B.7}$$

$$\hat{\lambda}^{n+1}(\omega) = \hat{\lambda}^n(\omega) + \tau \left(\hat{f}(\omega) - \sum_k \hat{u}_k^{n+1}(\omega) \right) \tag{B.8}$$

where τ denotes the noise tolerance. The above iterative calculations (Eqs. (B.6)–(B.8)) continue until the following convergence (Eq. (B.9)) is reached:

$$\frac{\sum_k \|\hat{u}_k^{n+1} - \hat{u}_k^n\|_2^2}{\sum_k \|\hat{u}_k^n\|_2^2} < \varepsilon \tag{B.9}$$

The final output will be the frequency spectrums $\hat{u}_k^{n+1}(\omega)$ of K mode components, which are then transformed into the time-domain signals by utilizing the inverse FT. The theory of VMD is briefly illustrated in this study as above; further, the details of the computation procedure can be found in Ref. [84]. The summarized version of the VMD algorithm is presented in Algorithm 1.

Algorithm 1 The process of VMD

Initialize: $\left\{\widehat{u}_k^{1}\right\}, \left\{\widehat{\omega}_k^{1}\right\}, \widehat{\lambda}^{1}, n \leftarrow 0$

Repeat
- $n \leftarrow n + 1$
- **for** $k = 1 : K$ **do**
  - update U_k for all $\omega \geq 0$ using Eq. (B.6)
  - update $\widehat{\omega}_k$ using Eq. (B.7)
- **end for**
- Update $\widehat{\lambda}^{n}(\omega)$ for all $\omega \geq 0$ using Eq. (B.8)

until convergence: $\frac{\sum_k \|\hat{u}_k^{n+1} - \hat{u}_k^n\|_2^2}{\sum_k \|\hat{u}_k^n\|_2^2} < \varepsilon$

obtain $u_k^{n+1}(t)$ by the fast Fourier transform of $\hat{u}_k^{n+1}(\omega)$
end

B.2. Theory of LSTM method

The following define the updating formula for the three-gate structure information, i.e., Eqs. (B.10)–(B.15).

$$f_t = \sigma\left(W_f \cdot \left[h_{t-1}, x_t\right] + b_f\right) \tag{B.10}$$

$$i_t = \sigma\left(W_i \cdot \left[h_{t-1}, x_t\right] + b_i\right) \tag{B.11}$$

$$C_t' = \tanh\left(W_C \cdot \left[h_{t-1}, x_t\right] + b_C\right) \tag{B.12}$$

$$C_t = f_t * C_{t-1} + i_t * C_t' \tag{B.13}$$

$$o_t = \sigma\left(W_0 \cdot \left[h_{t-1}, x_t\right] + b_0\right) \tag{B.14}$$

$$h_t = o_t * \tanh\left(C_t\right) \tag{B.15}$$

where W and b present the corresponding gate weights and deviation values; $[h_{t-1}, x_t]$ means to connect two vectors into a longer vector; x_t represents the input value; t represents the current moment; $t-1$ represents the previous moment; h represents hidden states; i_t is the output of the input gate at time t; C_t is the current memory cell state; f_t and o_t are the outputs of the forget gate and the output gate at the current moment; σ is the sigmoid function while tanh is the hyperbolic tangent function; C_t' is the candidate value added to the new cell state and h_t is the output matrix.

B.3. Theory of CNN-LSTM model

Each convolution layer has several convolution kernels. Following the completion of convolution calculations, the results of each layer's convolution are non-linearly processed using an activation function. Commonly used activation functions include the Sigmoid activation function, Tanh activation function, and ReLU activation function. This study utilizes the ReLU activation function to process the convolution results. While the convolution layer extracts data features, the resulting feature dimensions are typically very high. To address this issue and reduce the network training cost, a pooling layer is added after the convolution layer to reduce feature dimension. Each convolutional layer can be depicted as:

$$y_{ij}^k = \sigma((\omega^k \otimes x)_{ij} + b_k) \tag{B.16}$$

where $\otimes$ represents the convolution operation, ω^k and b_k represent the weight and deviation of the kth layer, respectively. Here the activation function $\sigma(x)$ is the Rectified Linear Unit ($ReLU$) function, expressed as:

$$\sigma(x) = \max(0, x) = \begin{cases} x_j, & x_j > 0 \\ 0, & x_j < 0 \end{cases}$$

where x is termed as input, x_j is input element and σ is $ReLU$ function. It is a nonlinear function that behaves like a linear one to learn the complex relationships of the input value (See Fig. B.15).

In the CLSTM model, an EarlyStopping(ES) step is applied to track the model's loss on a validation dataset as per Fig. B.16. If a validation loss did not decrease for at least 10 consecutive epochs, the model training was terminated with best model parameters obtained in training phase. After hyperparameters are optimized as per Table B.12,

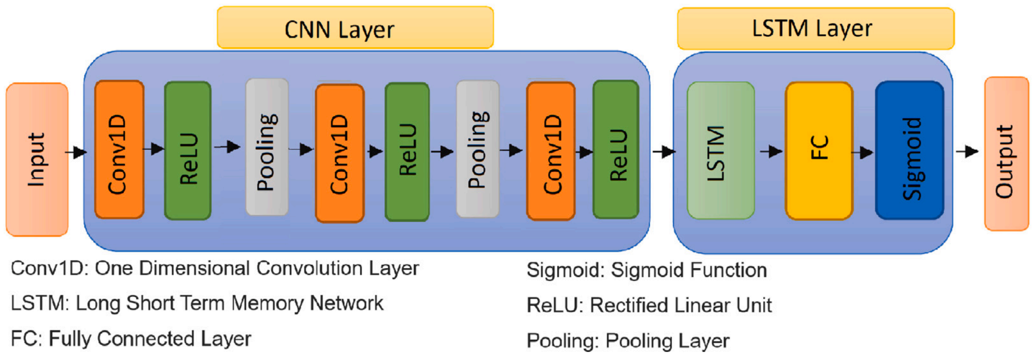


Fig. B.15. The block diagram of hybrid CNN-LSTM model.

Table B.12
Hyperparameters search range for the models using Bayesian optimization method. (*Note: 'gbtree' refers to the Gradient Boosted Trees, 'uniform' returns a random integer in the range [0, upper), 'quniform' means quantized version of 'uniform', quniform(low, high, q) return values in interval of 'q', ReLU is the Rectified Linear Unit Activation Function, and Adam is the optimizer known as Adaptive Moment Estimation.*)

Predictive models	Model hyperparameters	Hyperparameter selection
Convolution Neural Network Integrated with Long Short Term Memory Network (CLSTM)	Filter1 (CNN)	('Filter1', range(50,120,5))
	Filter 2 (CNN)	('Filter1', range(50,100,5))
	LSTM cell 1	('Units 1', range(50,100,5))
	LSTM cell 2	('Units 2', range(50,80,5))
	Epochs (CNN)	[1000]
	Activation function	[ReLU]
	Solver	['Adam']
	Batch size	('Batch_Size', range(50,1500,200))
Deep Neural Network (DNN)	Hiddenneuron 1	('Units2', range(50,150,5))
	Hiddenneuron 2	('Units3', range(50,80,5))
	Hiddenneuron 3	('Units4', range(50,50,5))
	Batch Size	('Batch_Size', range(50,1500,200))
	Solver	['Adam']
	Epochs	[1000]
	The maximum depth of the tree.	('max_depth', range(1,20,1))
Random Forest Regression (RF)	The number of trees in the forest.	('n_estimators', range(5,100,2))
	Minimum number of samples to split an internal node	('min_samples_split', range(2,100,1))
	The number of features to consider when looking for the best split.	['auto', 'sqrt', 'log2']
Long Short Term Memory Network (LSTM)	LSTM cell 1	('Units 1', range(50,100,5))
	LSTM cell 2	('Units 2', range(50,80,5))
	Activation function	[ReLU]
	Epochs	[1000]
	Drop rate	('drop_rate', range(0,0.5,0.1))
	Batch Size	('Batch_Size', range(50,1500,200))
eXtreme Gradient Boosting (XGB)	Booster Type	'gbtree'
	Step size shrinkage used in update to prevent overfitting.	('eta', range(0.1,0.9,0.1))
	The maximum depth of the tree.	('max_depth', range(1,20,1))
	The number of trees in the forest.	('n_estimators', range(5,100,2))
	Minimum sum of instance weight (hessian) needed in a child.	quniform('min_child_weight', 0, 10, 1)
	Parameters for subsampling of columns.	uniform('colsample_bytree', 0.5,1),
	L2 regularization term on weights	uniform('reg_lambda', 0,1)
	L1 regularization term on weights	uniform('reg_alpha', 0,1)
	Minimum loss reduction required to make a further partition on a leaf node of the tree.	uniform ('gamma', 1,9)

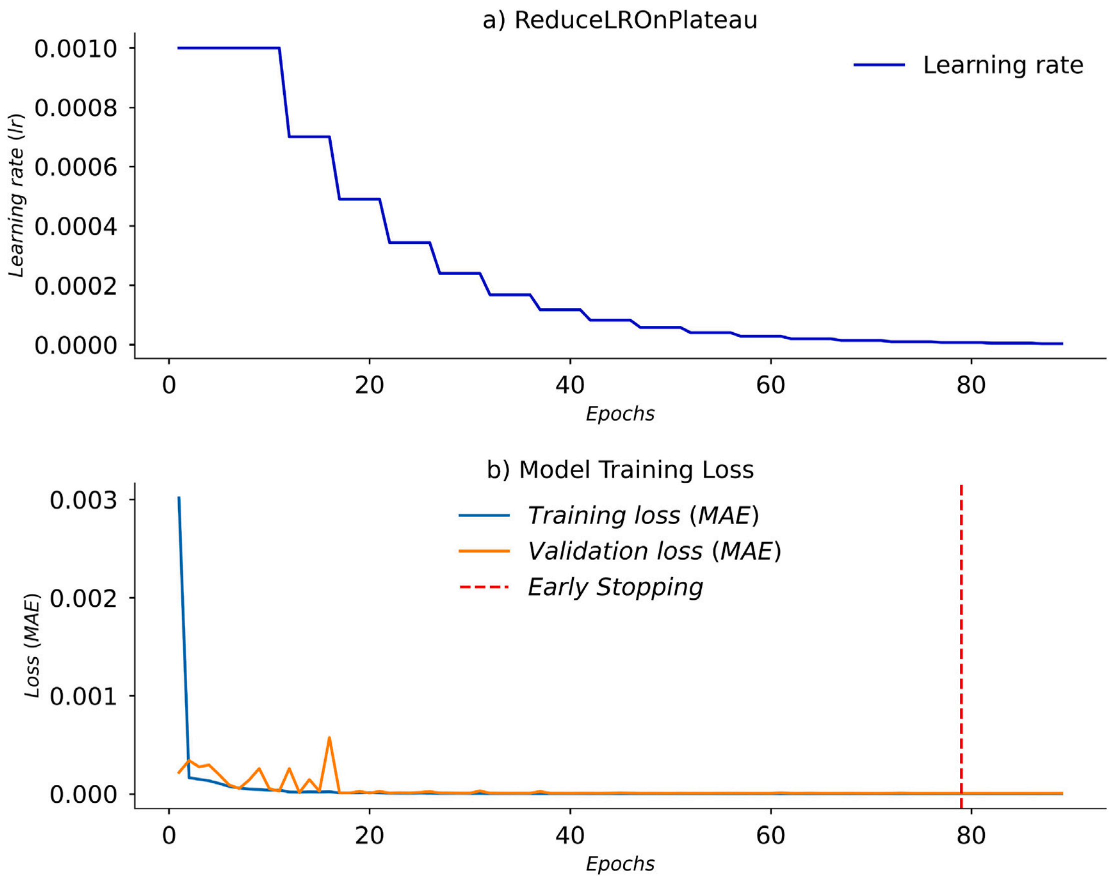


Fig. B.16. (a) Learning rate change in training using "ReduceLRonPlateau". (b) Loss or mean square error of both the training and validation datasets are calculated. Early stopping callbacks are implemented to stop the model from training further if there is no improvement in the validation loss for a specified number of epochs (*Figure valid only for VMF1 training of DS5)*.

```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 conv1d (Conv1D)             (None, 2, 115)            1265

 conv1d_1 (Conv1D)           (None, 2, 75)             69075

 lstm (LSTM)                 (None, 2, 100)            70400

 lstm_1 (LSTM)               (None, 50)                30200

 flatten (Flatten)           (None, 50)                0

 dense (Dense)               (None, 1)                 51

=================================================================
Total params: 170,991
Trainable params: 170,991
Non-trainable params: 0
_________________________________________________________________
```

Fig. B.17. The Keras model.summary() screenshot of the CLSTM, displays the network layers' output sizes and number of parameters. The total number of parameters is also indicated at the bottom of the image (*Figure valid only for VMF1 training of DS5)*.

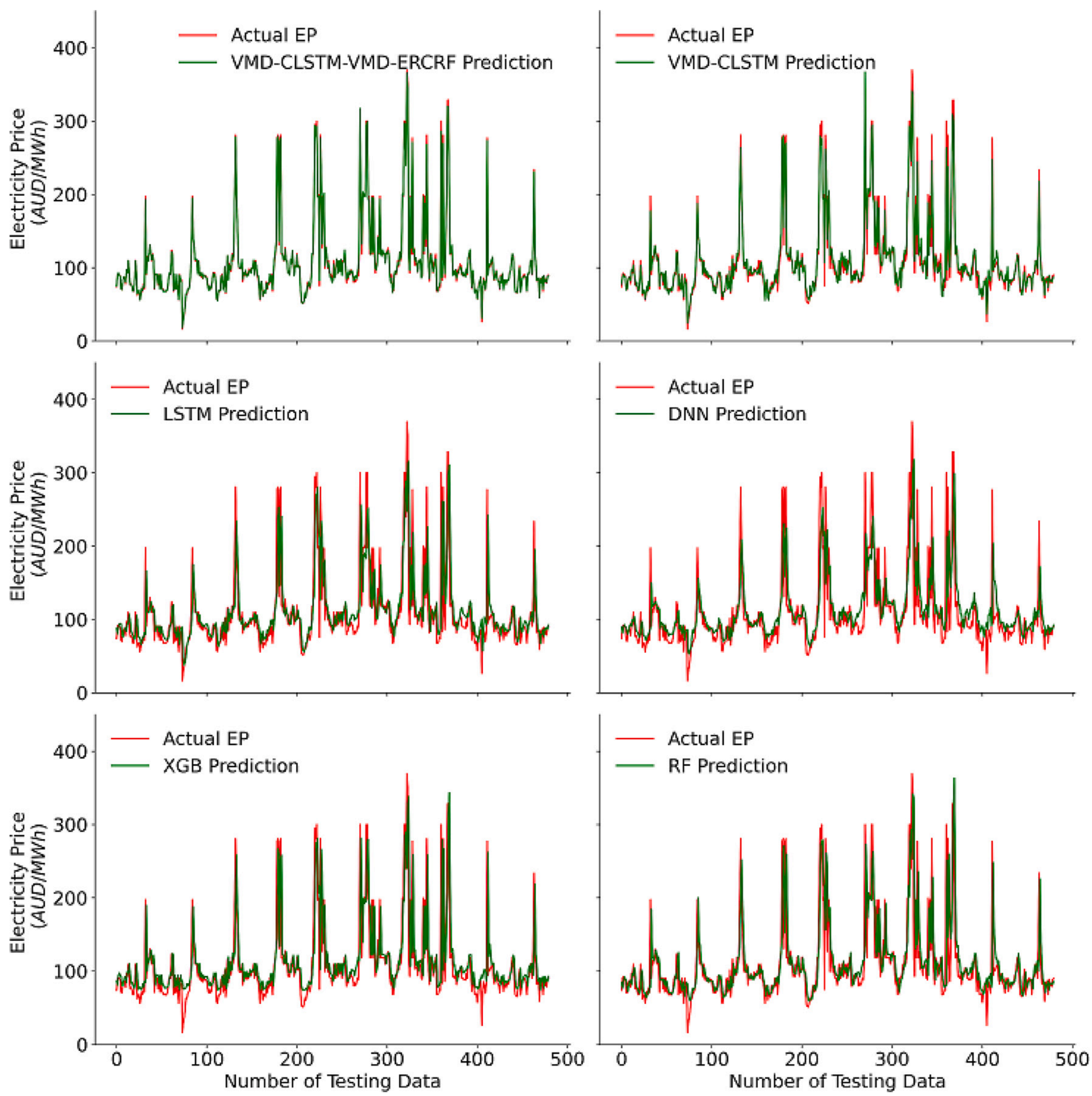


Fig. C.18. Actual *vs.* predicted half-hourly *EP* generated by the half-hourly VMD-CLSTM-VMD-ERCRF hybrid model in the testing phase, shown for 1-day dataset for the case of $DS2$. Comparison models are: (b) VMD-CLSTM, (c) LSTM, (d) DNN, (e) XGB and (f) RF. The relative error encountered is indicated in the blue colour.

the CLSTM model adopt Adaptive Moment Estimation (Adam) as a widely-used optimizer with "ReLU" as the activation function and the LSTM layers are succeeded by a recurrent dropout layer with dropout rate of 0.1. Fig. B.17 shows the parameters and network output size in training phase of VMF1 for DS5 in this study.

Appendix C. Supplementary results

Figs. C.18–C.21 show the actual *vs.* predicted half-hourly *EP* generated by the proposed VMD-CLSTM-VMD-ERCRF hybrid model for half-hourly *EP* predictions in the testing phase, albeit shown for 1-day dataset for the case of $DS2$, $DS3$, $DS4$ and $DS5$, respectively.

Similar to Tables 3 and 4, Table C.13 now evaluates the proposed VMD-CLSTM-VMD-ERCRF model for half-hourly EP prediction by also providing the CLSTM model without the data decomposition (i.e. VMD) method. Note that the CLSTM model is not reported elsewhere in the body of the paper for conciseness but here, the purpose is to demonstrate the influence of the VMD error correction strategy on CLSTM model's output. It is important to note that the application of VMD-based error correction strategy has led to a significant improvement in the model performance in terms of the Willmott's, Nash, and the

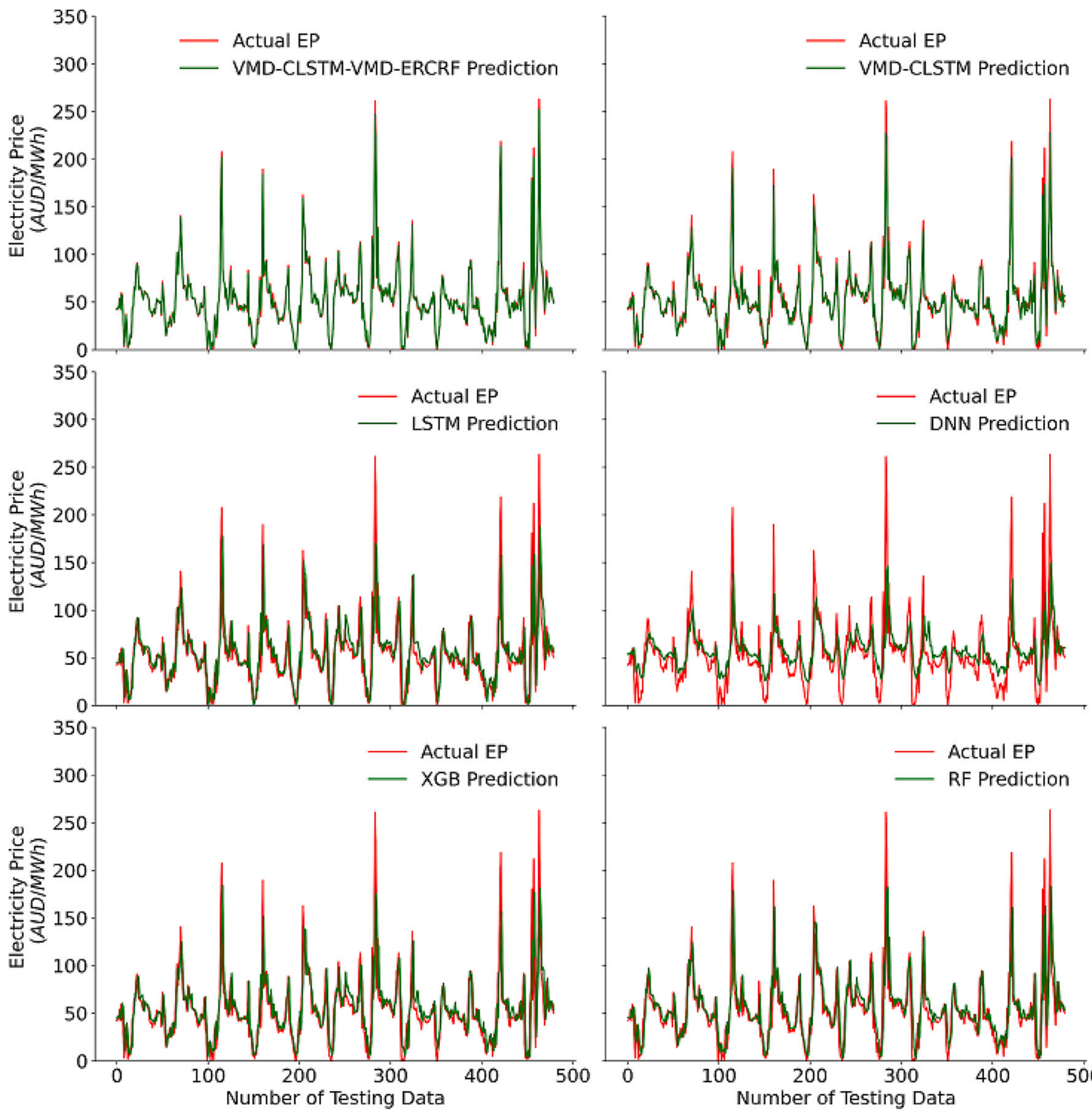


Fig. C.19. As per Fig. 9 but for *DS3*.

Legate's Indices, as well attaining as a lower error for the VMD-CLSTM compared to the CLSTM model. This reaffirms the efficacy of the variational mode decomposition algorithm and the error correction method in predicting half-hourly electricity demand.

To provide comprehensive insights into the predictive performance of different models from multiple perspectives, we have included Fig. C.22. This heatmap visually represents the $RMSE$ values for the objective model (i.e., VMD-CLSTM-VMD-ERCRF) in comparison to six benchmark models across all five datasets (DS1, DS2, DS3, DS4, and DS5). However, the simulations are performed at eight specific time intervals: 00:00 AM, 3:00 AM, 6:00 AM, 9:00 AM, 12:00 PM, 3:00 PM, 6:00 PM, and 9:00 PM. Furthermore, we have introduced the CLSTM (a standalone model) model for comparative analysis with the objective model (VMD-CLSTM-VMD-ERCRF), as well as VMD-CLSTM (a single decomposition model).

Upon examining the results of the objective model in contrast to other benchmark models across all five datasets, it becomes evident that VMD-CLSTM-VMD-ERCRF consistently yields predictions with lower $RMSE$ scores when compared to VMD-CLSTM, CLSTM, LSTM, DNN, XGB, and RF models. The figure reveals that the VMD-CLSTM-VMD-ERCRF model closely aligns with EP values at all time points under consideration. The elevated $RMSE$ values observed for DNN, LSTM, CLSTM, XGB, and RF models underscore their limitations in providing a suitable alternative for EP prediction. On the contrary, the lower $RMSE$ values for the VMD-CLSTM model compared to CLSTM, LSTM, DNN, XGB, and RF models highlight that a single model may not be sufficient for accurate predictions in scenarios involving high volatility, nonstationary, multi-seasonality, and nonlinearity in EP, thereby emphasizing the significant value offered by decomposition models in prediction.

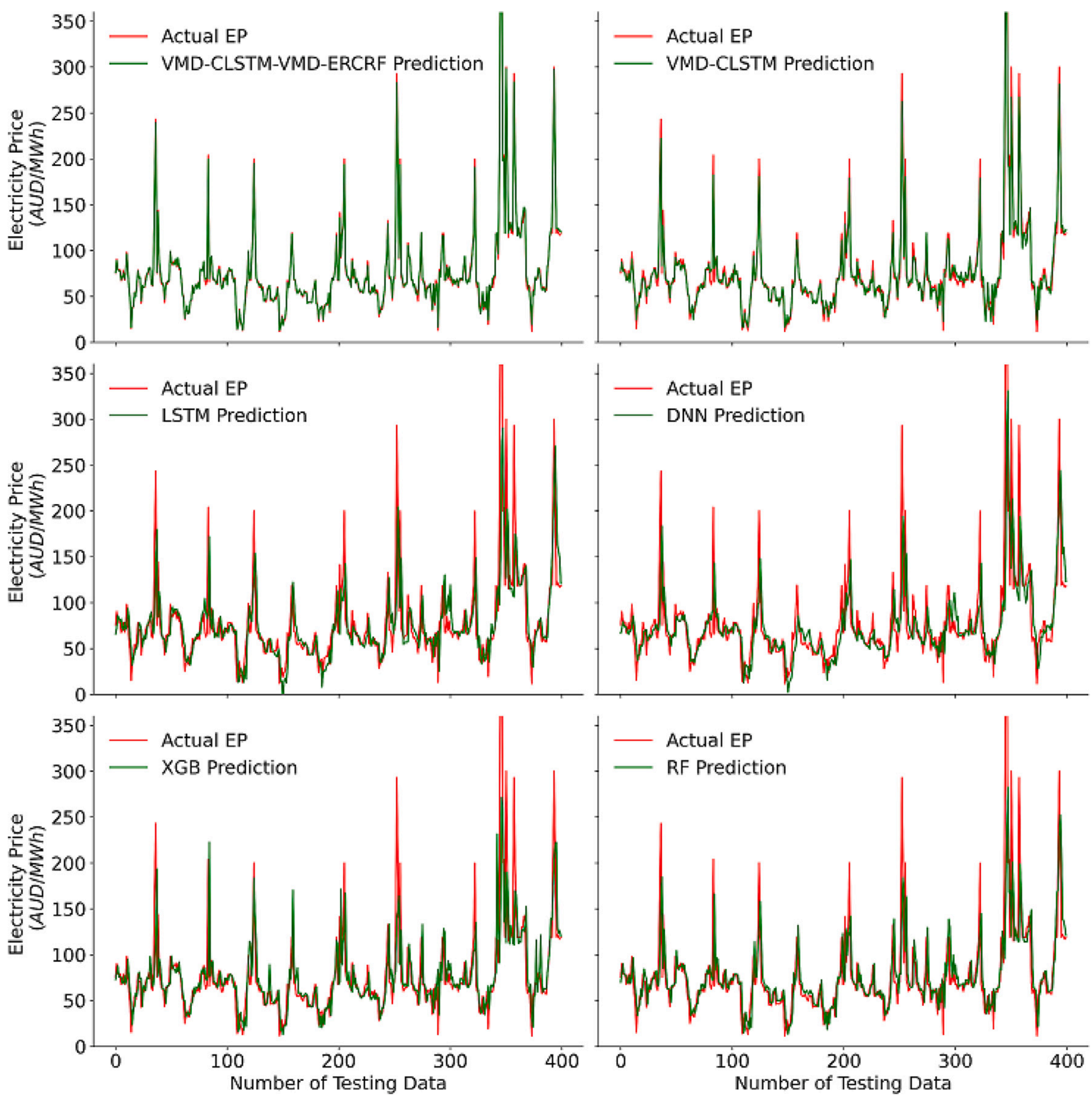


Fig. C.20. As per Fig. 9 but for *DS4*.

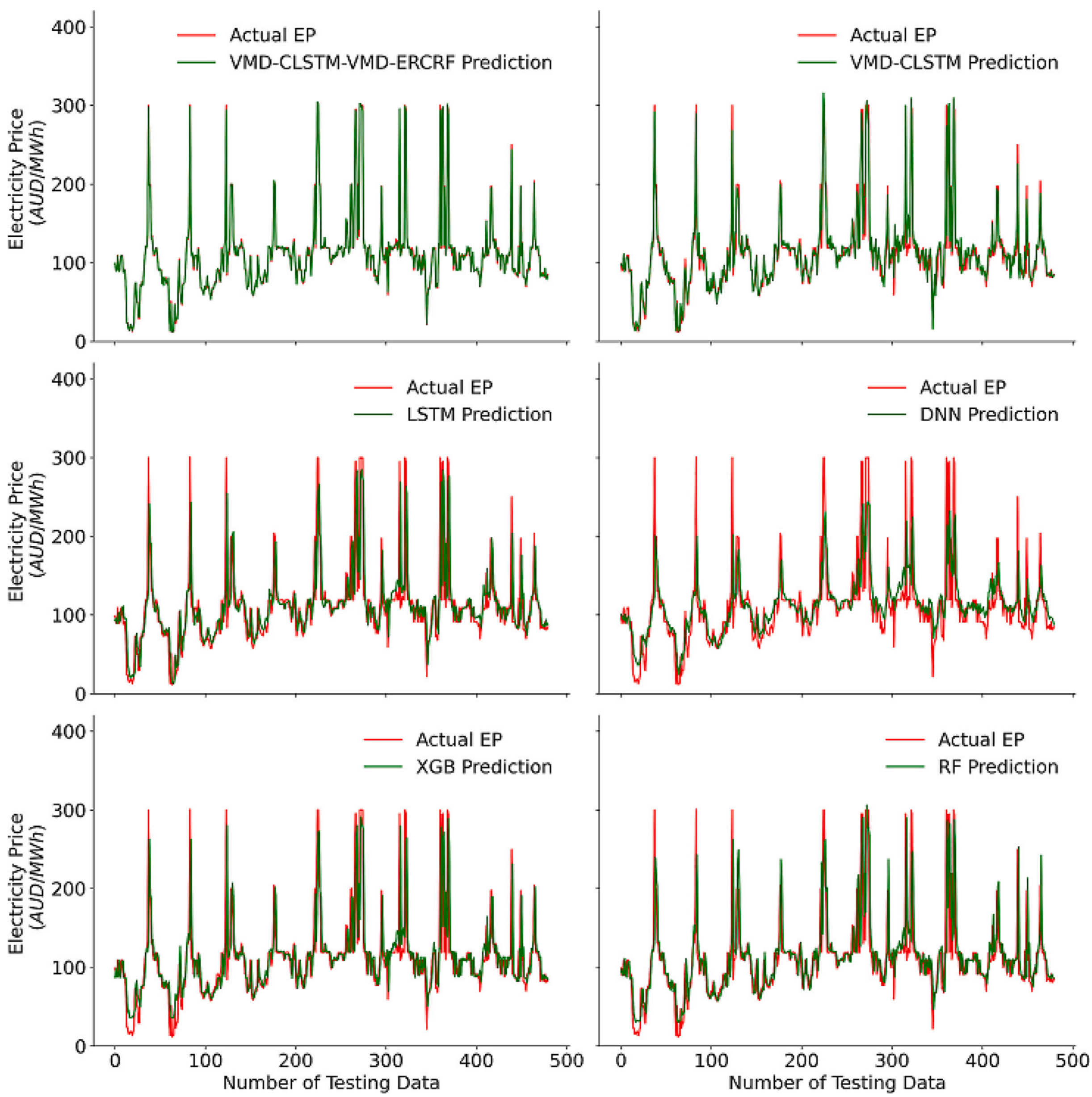


Fig. C.21. As per Fig. 9 but for *DS5*.

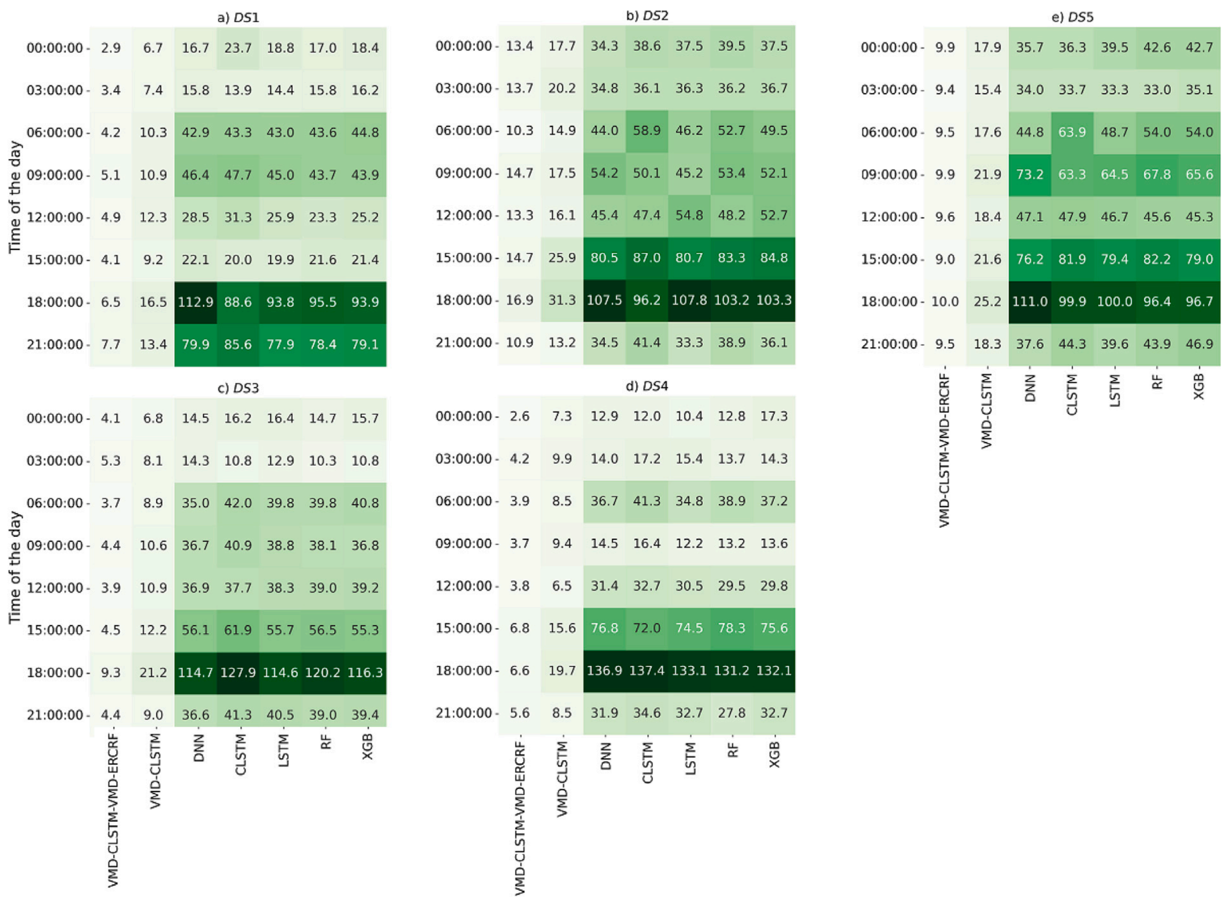


Fig. C.22. A heatmap visually representing the $RMSE$ (Root Mean Square Error) values for the Objective model in comparison to benchmark models across 3-h intervals for five datasets. The colour scheme represents $RMSE$ values in AUD/MWh, highlighting discrepancies between the models. The colour gradient shifts from green (indicating lower $RMSE$) to dark green (indicating higher $RMSE$). Each cell displays the specific $RMSE$ values for each of the models.

Table C.13
Evaluating the proposed VMD-CLSTM-VMD-ERCRF model for half-hourly *EP* predictions. Note: R^2 = Coefficient of Determination, $RMSE$: AUD/MWh = Root Mean Square Error; MAE:*AUD/MWh* = Mean Absolute Error; I_{WI} = Wilmott's Index; I_{NS} = Nash–Sutcliffe Index; I_{LM} = Legates and McCabe Index; APB = Absolute percentage Bias. The objective model is benchmarked against the VMD-CLSTM, CLSTM, XGB, RF, DNN and the LSTM models for four different seasons and the yearly (i.e., 2022) prediction dataset with the best model indicated in blue.

Dataset	Predictive models	R^2	I_{WI}	I_{NS}	I_{LM}	RMSE	MAE	APB
DS1	**VMD-CLSTM-VMD-ERCRF**	**0.999**	**0.994**	**0.995**	**0.935**	**5.191**	**3.057**	**3.66%**
	VMD-CLSTM	0.874	0.860	0.845	0.785	11.352	6.770	8.11%
	CLSTM	0.877	0.605	0.584	0.481	46.265	24.313	29.14%
	LSTM	0.882	0.644	0.602	0.514	45.171	22.781	27.31%
	DNN	0.858	0.548	0.539	0.456	48.574	25.486	30.55%
	XGB	0.883	0.625	0.607	0.511	44.846	22.902	27.45%
	RF	0.882	0.638	0.605	0.508	44.983	23.050	27.63%
DS2	**VMD-CLSTM-VMD-ERCRF**	**0.998**	**0.992**	**0.988**	**0.916**	**13.190**	**8.250**	**4.27%**
	VMD-CLSTM	0.916	0.882	0.873	0.767	19.412	13.124	6.79%
	CLSTM	0.929	0.828	0.742	0.628	60.453	36.685	18.97%
	LSTM	0.911	0.839	0.752	0.648	59.306	34.654	17.93%
	DNN	0.929	0.825	0.744	0.634	60.199	36.105	18.68%
	XGB	0.897	0.817	0.738	0.631	61.038	36.375	18.82%
	RF	0.930	0.826	0.746	0.639	60.123	35.595	18.41%
DS3	**VMD-CLSTM-VMD-ERCRF**	**0.998**	**0.983**	**0.990**	**0.897**	**6.247**	**3.388**	**4.55%**
	VMD-CLSTM	0.902	0.832	0.862	0.716	12.333	7.062	9.49%
	CLSTM	0.747	0.369	0.271	0.238	54.017	25.123	33.73%
	LSTM	0.776	0.377	0.358	0.266	50.652	24.178	32.48%
	DNN	0.779	0.298	0.362	0.250	50.510	24.704	33.18%
	XGB	0.770	0.369	0.351	0.277	50.944	23.827	32.00%
	RF	0.770	0.364	0.350	0.286	50.957	23.530	31.61%
DS4	**VMD-CLSTM-VMD-ERCRF**	**0.999**	**0.992**	**0.995**	**0.930**	**4.748**	**2.507**	**2.55%**
	VMD-CLSTM	0.891	0.877	0.882	0.835	11.367	5.900	6.00%
	CLSTM	0.821	0.478	0.453	0.401	49.870	21.375	21.71%
	LSTM	0.832	0.513	0.477	0.415	48.753	20.855	21.19%
	DNN	0.817	0.489	0.429	0.398	50.921	21.467	21.81%
	XGB	0.839	0.512	0.494	0.433	47.923	20.223	20.55%
	RF	0.837	0.506	0.489	0.453	48.168	19.497	19.81%
DS5	**VMD-CLSTM-VMD-ERCRF**	**0.999**	**0.996**	**0.994**	**0.953**	**10.220**	**5.154**	**2.61%**
	VMD-CLSTM	0.914	0.858	0.877	0.819	21.855	13.178	6.67%
	CLSTM	0.747	0.369	0.271	0.238	54.017	25.123	33.73%
	LSTM	0.906	0.851	0.801	0.680	61.125	34.886	17.66%
	DNN	0.872	0.828	0.786	0.654	63.370	37.680	19.07%
	XGB	0.926	0.845	0.798	0.672	61.584	35.690	18.06%
	RF	0.896	0.844	0.798	0.672	61.651	35.692	18.06%

References

[1] Wang K, Yu M, Niu D, Liang Y, Peng S, Xu X. Short-term electricity price forecasting based on similarity day screening, two-layer decomposition technique and Bi-LSTM neural network. Appl Soft Comput 2023;110018.

[2] Zhang T, Tang Z, Wu J, Du X, Chen K. Short term electricity price forecasting using a new hybrid model based on two-layer decomposition technique and ensemble learning. Electr Power Syst Res 2022;205:107762.

[3] Lin K-P, Pai P-F, Yang S-L. Forecasting concentrations of air pollutants by logarithm support vector regression with immune algorithms. Appl Math Comput 2011;217(12):5318–27.

[4] Mandal P, Senjyu T, Funabashi T. Neural networks approach to forecast several hour ahead electricity prices and loads in deregulated market. Energy Convers Manage 2006;47(15–16):2128–42.

[5] Al-Musaylh MS, Deo RC, Li Y, Adamowski JF. Two-phase particle swarm optimized-support vector regression hybrid model integrated with improved empirical mode decomposition with adaptive noise for multiple-horizon electricity demand forecasting. Appl Energy 2018;217:422–39.

[6] Wang J, Zhang L, Li Z. Interval forecasting system for electricity load based on data pre-processing strategy and multi-objective optimization algorithm. Appl Energy 2022;305:117911.

[7] Tan YQ, Shen YX, Yu XY, Lu X. Day-ahead electricity price forecasting employing a novel hybrid frame of deep learning methods: A case study in NSW, Australia. Electr Power Syst Res 2023;220:109300.

[8] Chen Y, Wang Y, Ma J, Jin Q. BRIM: An accurate electricity spot price prediction scheme-based bidirectional recurrent neural network and integrated market. Energies 2019;12(12):2241.

[9] Ghimire S, Yaseen ZM, Farooque AA, Deo RC, Zhang J, Tao X. Streamflow prediction using an integrated methodology based on convolutional neural network and long short-term memory networks. Sci Rep 2021;11(1):17497.

[10] Jayasinghe WLP, Deo RC, Ghahramani A, Ghimire S, Raj N. Deep multi-stage reference evapotranspiration forecasting model: Multivariate empirical mode decomposition integrated with the boruta-random forest algorithm. IEEE Access 2021;9:166695–708.

[11] Jayasinghe WLP, Deo RC, Ghahramani A, Ghimire S, Raj N. Development and evaluation of hybrid deep learning long short-term memory network model for pan evaporation estimation trained with satellite and ground-based data. J Hydrol 2022;607:127534.

[12] Zhang J, Tan Z, Wei Y. An adaptive hybrid model for short term electricity price forecasting. Appl Energy 2020;258:114087.

[13] Yang H, Schell KR. QCAE: A quadruple branch CNN autoencoder for real-time electricity price forecasting. Int J Electr Power Energy Syst 2022;141:108092.

[14] Ghimire S, Deo RC, Wang H, Al-Musaylh MS, Casillas-Pérez D, Salcedo-Sanz S. Stacked LSTM sequence-to-sequence autoencoder with feature selection for daily solar radiation prediction: a review and new modeling results. Energies 2022;15(3):1061.

[15] Ghimire S, Deo RC, Casillas-Pérez D, Salcedo-Sanz S, Sharma E, Ali M. Deep learning CNN-LSTM-MLP hybrid fusion model for feature optimizations and daily solar radiation prediction. Measurement 2022;111759.

[16] Deo RC, Grant RH, Webb A, Ghimire S, Igoe DP, Downs NJ, Al-Musaylh MS, Parisi AV, Soar J. Forecasting solar photosynthetic photon flux density under cloud cover effects: novel predictive model using convolutional neural network integrated with long short-term memory network. In: Stochastic environmental research and risk assessment. Springer; 2022, p. 1–38.

[17] Ghimire S, Deo RC, Casillas-Pérez D, Salcedo-Sanz S. Boosting solar radiation predictions with global climate models, observational predictors and hybrid deep-machine learning algorithms. Appl Energy 2022;316:119063.

[18] Lago J, De Ridder F, De Schutter B. Forecasting spot electricity prices: Deep learning approaches and empirical comparison of traditional algorithms. Appl Energy 2018;221:386–405.

[19] Gökgöz F, Filiz F. Electricity price forecasting: A comparative analysis with shallow-ann and dnn. Int J Energy Power Eng 2018;12(6):421–5.

[20] Wang L, Zhang Z, Chen J. Short-term electricity price forecasting with stacked denoising autoencoders. IEEE Trans Power Syst 2016;32(4):2673–81.

[21] Wang S, Wang J, Lu H, Zhao W. A novel combined model for wind speed prediction–Combination of linear model, shallow neural networks, and deep learning approaches. Energy 2021;234:121275.

[22] Xiong X, Qing G. A hybrid day-ahead electricity price forecasting framework based on time series. Energy 2023;264:126099.

[23] Wang Y, Wang J, Li Z, Yang H, Li H. Design of a combined system based on two-stage data preprocessing and multi-objective optimization for wind speed prediction. Energy 2021;231:121125.
[24] Saâdaoui F, Rabbouch H. A wavelet-based hybrid neural network for short-term electricity prices forecasting. Artif Intell Rev 2019;52:649–69.
[25] Yang Z, Ce L, Lian L. Electricity price forecasting by a hybrid model, combining wavelet transform, ARMA and kernel-based extreme learning machine methods. Appl Energy 2017;190:291–305.
[26] Yan X, Chowdhury NA. Mid-term electricity market clearing price forecasting utilizing hybrid support vector machine and auto-regressive moving average with external input. Int J Electr Power Energy Syst 2014;63:64–70.
[27] Shang J, Gao J, Jiang X, Liu M, Liu D. Optimal configuration of hybrid energy systems considering power to hydrogen and electricity-price prediction: A two-stage multi-objective bi-level framework. Energy 2023;263:126023.
[28] Zhang C, Fu Y. Probabilistic electricity price forecast with optimal prediction interval. IEEE Trans Power Syst 2023.
[29] Billé AG, Gianfreda A, Del Grosso F, Ravazzolo F. Forecasting electricity prices with expert, linear, and nonlinear models. Int J Forecast 2023;39(2):570–86.
[30] Pandey A, Pandey M. A robust neural network based short time electricity price prediction. In: 2023 3rd International conference on intelligent communication and computational techniques (ICCT). IEEE; 2023, p. 1–8.
[31] Abdellatif A, Mubarak H, Ahmad S, Mekhilef S, Abdellatef H, Mokhlis H, Kanesan J. Electricity price forecasting one day ahead by employing hybrid deep learning model. In: 2023 IEEE IAS global conference on renewable energy and hydrogen technologies (GlobConHT). IEEE; 2023, p. 1–5.
[32] Zhang B, Song C, Jiang X, Li Y. Electricity price forecast based on the STL-TCN-NBEATS model. Heliyon 2023;e13029.
[33] Xu F, Teng X, Lu J, Zheng T, Jin Y. Prediction of day-ahead electricity price based on N-BEATSx model optimized by SSA considering coupling between features. In: Proceedings of the 7th PURPLE MOUNTAIN FORUM on smart grid protection and control (PMF2022). Springer; 2023, p. 178–94.
[34] Mubarak H, Ahmad S, Hossain AA, Horan B, Abdellatif A, Mekhilef S, Seyedmahmoudian M, Stojcevski A, Mokhlis H, Kanesan J, et al. Short-term electricity price forecasting using interpretable hybrid machine learning models. In: 2023 IEEE IAS global conference on renewable energy and hydrogen technologies (GlobConHT). IEEE; 2023, p. 1–6.
[35] Zhang Y, Tao P, Wu X, Yang C, Han G, Zhou H, Hu Y. Hourly electricity price prediction for electricity market with high proportion of wind and solar power. Energies 2022;15(4):1345.
[36] Gabrielli P, Wüthrich M, Blume S, Sansavini G. Data-driven modeling for long-term electricity price forecasting. Energy 2022;244:123107.
[37] Arvanitidis AI, Bargiotas D, Kontogiannis D, Fevgas A, Alamaniotis M. Optimized data-driven models for short-term electricity price forecasting based on signal decomposition and clustering techniques. Energies 2022;15(21):7929.
[38] Kontogiannis D, Bargiotas D, Daskalopulu A, Arvanitidis AI, Tsoukalas LH. Error compensation enhanced day-ahead electricity price forecasting. Energies 2022;15(4):1466.
[39] McHugh C, Coleman S, Kerr D. Hourly electricity price forecasting with NARMAX. Mach Learn Appl 2022;9:100383.
[40] Sridharan V, Tuo M, Li X. Wholesale electricity price forecasting using integrated long-term recurrent convolutional network model. Energies 2022;15(20):7606.
[41] Li W, Becker DM. Day-ahead electricity price prediction applying hybrid models of LSTM-based deep learning methods and feature selection algorithms under consideration of market coupling. Energy 2021;237:121543.
[42] Yang H, Schell KR. Real-time electricity price forecasting of wind farms with deep neural network transfer learning and hybrid datasets. Appl Energy 2021;299:117242.
[43] Chen X, Dong ZY, Meng K, Xu Y, Wong KP, Ngan H. Electricity price forecasting with extreme learning machine and bootstrapping. IEEE Trans Power Syst 2012;27(4):2055–62.
[44] Brusaferri A, Matteucci M, Portolani P, Vitali A. Bayesian deep learning based method for probabilistic forecast of day-ahead electricity prices. Appl Energy 2019;250:1158–75.
[45] Ugurlu U, Oksuz I, Tas O. Electricity price forecasting using recurrent neural networks. Energies 2018;11(5):1255.
[46] Dedinec A, Filiposka S, Dedinec A, Kocarev L. Deep belief network based electricity load forecasting: An analysis of Macedonian case. Energy 2016;115:1688–700.
[47] Zhang J, Tan Z, Li C. A novel hybrid forecasting method using GRNN combined with wavelet transform and a GARCH model. Energy Sources B 2015;10(4):418–26.
[48] Ludwig N, Feuerriegel S, Neumann D. Putting Big Data analytics to work: Feature selection for forecasting electricity prices using the LASSO and random forests. J Decis Syst 2015;24(1):19–36.
[49] Chaâbane N. A novel auto-regressive fractionally integrated moving average–least-squares support vector machine model for electricity spot prices prediction. J Appl Stat 2014;41(3):635–51.
[50] Lin W-M, Gow H-J, Tsai M-T. An enhanced radial basis function network for short-term electricity price forecasting. Appl Energy 2010;87(10):3226–34.
[51] Amjady N, Keynia F. Electricity market price spike analysis by a hybrid data model and feature selection technique. Electr Power Syst Res 2010;80(3):318–27.
[52] Catalão JPdS, Pousinho HMI, Mendes VMF. Hybrid wavelet-PSO-ANFIS approach for short-term electricity prices forecasting. IEEE Trans Power Syst 2010;26(1):137–44.
[53] Che J, Wang J. Short-term electricity prices forecasting based on support vector regression and auto-regressive integrated moving average modeling. Energy Convers Manage 2010;51(10):1911–7.
[54] Pindoriya N, Singh S, Singh S. An adaptive wavelet neural network-based energy price forecasting in electricity markets. IEEE Trans Power Syst 2008;23(3):1423–32.
[55] Yamin H, Shahidehpour S, Li Z. Adaptive short-term electricity price forecasting using artificial neural networks in the restructured power markets. Int J Electr Power Energy Syst 2004;26(8):571–81.
[56] Szkuta BR, Sanabria LA, Dillon TS. Electricity price short-term forecasting using artificial neural networks. IEEE Trans Power Syst 1999;14(3):851–7.
[57] Wan C, Xu Z, Wang Y, Dong ZY, Wong KP. A hybrid approach for probabilistic forecasting of electricity price. IEEE Trans Smart Grid 2013;5(1):463–70.
[58] Memarzadeh G, Keynia F. Short-term electricity load and price forecasting by a new optimal LSTM-NN based prediction algorithm. Electr Power Syst Res 2021;192:106995.
[59] Xie X, Xu W, Tan H. The day-ahead electricity price forecasting based on stacked CNN and LSTM. In: Intelligence science and big data engineering: 8th International conference, IScIDE 2018, Lanzhou, China, August 18–19, 2018, Revised selected papers 8. Springer; 2018, p. 216–30.
[60] Ghimire S, Nguyen-Huy T, AL-Musaylh MS, Deo RC, Casillas-Pérez D, Salcedo-Sanz S. A novel approach based on integration of convolutional neural networks and echo state network for daily electricity demand prediction. Energy 2023;127430.
[61] Kuo P-H, Huang C-J. An electricity price forecasting model by hybrid structured deep neural networks. Sustainability 2018;10(4):1280.
[62] Heidarpanah M, Hooshyaripor F, Fazeli M. Daily electricity price forecasting using artificial intelligence models in the Iranian electricity market. Energy 2023;263:126011.
[63] Osório GJ, Lotfi M, Shafie-Khah M, Campos VM, Catalão JP. Hybrid forecasting model for short-term electricity market prices with renewable integration. Sustainability 2018;11(1):57.
[64] He K, Wang H, Du J, Zou Y. Forecasting electricity market risk using empirical mode decomposition (EMD)—based multiscale methodology. Energies 2016;9(11):931.
[65] Rani R, Jessie H, Victoire T, Albert A. A hybrid Elman recurrent neural network, group search optimization, and refined VMD-based framework for multi-step ahead electricity price forecasting. Soft Comput 2019;23(18):8413–34.
[66] Zhang H, Yang Y, Zhang Y, He Z, Yuan W, Yang Y, Qiu W, Li L. A combined model based on SSA, neural networks, and LSSVM for short-term electric load and price forecasting. Neural Comput Appl 2021;33:773–88.
[67] Azam MF, Younis MS. Multi-horizon electricity load and price forecasting using an interpretable multi-head self-attention and EEMD-based framework. IEEE Access 2021;9:85918–32.
[68] Zhang J, Li D, Hao Y, Tan Z. A hybrid model using signal processing technology, econometric models and neural network for carbon spot price forecasting. J Clean Prod 2018;204:958–64.
[69] Yang W, Wang J, Niu T, Du P. A hybrid forecasting system based on a dual decomposition strategy and multi-objective optimization for electricity price forecasting. Appl Energy 2019;235:1205–25.
[70] Ghimire S, Deo RC, Casillas-Pérez D, Salcedo-Sanz S. Improved complete ensemble empirical mode decomposition with adaptive noise deep residual model for short-term multi-step solar radiation prediction. Renew Energy 2022;190:408–24.
[71] Qiao W, Yang Z. Forecast the electricity price of US using a wavelet transform-based hybrid model. Energy 2020;193:116704.
[72] Conejo AJ, Plazas MA, Espinola R, Molina AB. Day-ahead electricity price forecasting using the wavelet transform and ARIMA models. IEEE Trans Power Syst 2005;20(2):1035–42.
[73] Huang C-J, Shen Y, Chen Y-H, Chen H-C. A novel hybrid deep neural network model for short-term electricity price forecasting. Int J Energy Res 2021;45(2):2511–32.
[74] Wang D, Luo H, Grunder O, Lin Y, Guo H. Multi-step ahead electricity price forecasting using a hybrid model based on two-layer decomposition technique and BP neural network optimized by firefly algorithm. Appl Energy 2017;190:390–407.
[75] Chang Z, Zhang Y, Chen W. Electricity price prediction based on hybrid model of adam optimized LSTM neural network and wavelet transform. Energy 2019;187:115804.
[76] Peng L, Liu S, Liu R, Wang L. Effective long short-term memory with differential evolution algorithm for electricity price prediction. Energy 2018;162:1301–14.
[77] Babu CN, Reddy BE. A moving-average filter based hybrid ARIMA–ANN model for forecasting time series data. Appl Soft Comput 2014;23:27–38.
[78] Alickovic E, Kevric J, Subasi A. Performance evaluation of empirical mode decomposition, discrete wavelet transform, and wavelet packed decomposition for automated epileptic seizure detection and prediction. Biomed Signal Process Control 2018;39:94–102.
[79] Forrest S, MacGill I. Assessing the impact of wind generation on wholesale prices and generator dispatch in the Australian National Electricity Market. Energy Policy 2013;59:120–32.

[80] Amjady N, Keynia F. A new prediction strategy for price spike forecasting of day-ahead electricity markets. Appl Soft Comput 2011;11(6):4246–56.
[81] Ghimire S, Nguyen-Huy T, Prasad R, Deo RC, Casillas-Pérez D, Salcedo-Sanz S, Bhandari B. Hybrid convolutional neural network-multilayer perceptron model for solar radiation prediction. Cogn Comput 2022;1–27.
[82] Castillo-Botón C, Casillas-Pérez D, Casanova-Mateo C, Ghimire S, Cerro-Prada E, Gutierrez P, Deo R, Salcedo-Sanz S. Machine learning regression and classification methods for fog events prediction. Atmos Res 2022;272:106157.
[83] Ghimire S, Nguyen-Huy T, Deo RC, Casillas-Perez D, Salcedo-Sanz S. Efficient daily solar radiation prediction with deep learning 4-phase convolutional neural network, dual stage stacked regression and support vector machine CNN-REGST hybrid model. Sustain Mater Technol 2022;32:e00429.
[84] Dragomiretskiy K, Zosso D. Variational mode decomposition. IEEE Trans Signal Process 2013;62(3):531–44.
[85] Chollet F, et al. Keras: Deep learning library for theano and tensorflow. 2015;7:T1. URL: https://keras.io/k.
[86] Močkus J. On Bayesian methods for seeking the extremum. In: Optimization techniques IFIP technical conference: Novosibirsk, July 1–7, 1974. Springer; 1975, p. 400–4.
[87] Ao Y, Li H, Zhu L, Ali S, Yang Z. The linear random forest algorithm and its advantages in machine learning assisted logging regression modeling. J Pet Sci Eng 2019;174:776–89.
[88] Despotovic M, Nedic V, Despotovic D, Cvetanovic S. Review and statistical analysis of different global solar radiation sunshine models. Renew Sustain Energy Rev 2015;52:1869–80.
[89] Diebold FX. Comparing predictive accuracy, twenty years later: A personal perspective on the use and abuse of Diebold–Mariano tests. J Bus Econom Statist 2015;33(1):1.
[90] Prasad SS, Deo RC, Downs NJ, Casillas-Pérez D, Salcedo-Sanz S, Parisi AV. Very short-term solar ultraviolet-A radiation forecasting system with cloud cover images and a Bayesian optimized interpretable artificial intelligence model. Expert Syst Appl 2023;121273.
[91] Prasad SS, Deo RC, Salcedo-Sanz S, Downs NJ, Casillas-Pérez D, Parisi AV. Enhanced joint hybrid deep neural network explainable artificial intelligence model for 1-hr ahead solar ultraviolet index prediction. Comput Methods Programs Biomed 2023;107737.
[92] Mwampashi MM, Sklibosios Nikitopoulos C, Rai A. From 30-to 5-minute settlement rule in the NEM: An early evaluation. 2022, Available at SSRN.
[93] AEMC. Five minute settlement, final determination, 28 November. Sydney: Australian Energy Market Commission.