



Short-term smart grid energy forecasting using a hybrid deep learning method on univariate and multivariate data sets

William Gomez ^{a,b} , Fu-Kwun Wang ^{a,*} , Shey-Huei Sheu ^{b,c}

^a Department of Industrial Management, National Taiwan University of Science and Technology, Taipei City, Taiwan

^b Department of Business Administration, Asia University, Taichung, Taiwan

^c Department of Medical Research, China Medical University Hospital, China Medical University, Taichung, Taiwan

ARTICLE INFO

Keywords:

Short-term energy forecasting
Deep neural network
Bidirectional long short term with attention
Uncertainty estimation
Smart grid

ABSTRACT

Transforming energy power systems into integrated smart grids is pivotal in addressing climate change challenges. Accurate forecasting of energy demand within such grids is essential for optimizing operations, incorporating renewable energy sources, and enhancing grid stability. However, many traditional or simpler machine learning approaches struggle to harness the interdependencies between the various energy sources. This limits their predictive performance and scalability. This paper presents a novel hybrid approach for short-term smart grid energy forecasting in the load and wind speed datasets, utilizing meteorological factors to enhance prediction accuracy. To reduce the difficulty caused by energy volatility, noisy measurement, non-linearity, and improve short-term forecasting, the proposed method combines an optimal intrinsic mode function decomposition method, a bidirectional long short-term memory (BiLSTM) model with an attention mechanism and integrating deep neural network (DNN) layers, while employing Bayesian optimization with Gaussian processes for hyperparameter tuning. The model effectively captures complex temporal and spatial dependencies in the univariate and multivariate data with different time horizons. Results demonstrate significant improvements in predictive accuracy, with the proposed model outperforming conventional single models and hybrid approaches across all forecast horizons. The proposed model also achieves a significant improvement in computational efficiency compared to the default decomposition method. Achieving an average reduction in computational time of 65.78 % for load data and 49.23 % for wind speed data. An improvement rate of up to 34.17 % for two-day forecasts and 5.37 % for seven-day forecasts for load data in terms of root mean square error. The findings suggest that the proposed method can be a reliable tool for smart grid operators and energy planners. This will aid decision-making processes related to grid operation and renewable energy integration.

1. Introduction

With the increasing advancement of smart industrialization and technology, energy consumption has risen significantly in the 21st century. Development and integration of smart grids and renewable energy sources have made substantial progress in many countries. The integrated energy system (IES) or smart grid has become a key driver in achieving net zero carbon emissions by 2050 [1,2]. Rapid advancements in smart manufacturing and energy evolution challenges, such as scheduling inefficiencies, pricing risks, overgeneration, and environmental issues, are intensified by rising grid usage and the growth of renewable energy generators, especially when supply and demand estimations are inaccurate. Technological progress and economic growth

have significantly increased electricity demand, making accurate short-to medium-term load forecasting crucial for the energy market. Electric utilities rely on the accuracy of these forecasts to synchronize electricity generation with consumption, which is vital for meeting energy needs, supporting clean energy initiatives, and fostering sustainable development [1,3]. Forecasting periods have different impacts on the energy management system (EMS). Developing flexible, robust, and adaptable forecasting models for grid systems is essential in response to the rapidly changing energy market. Recent literature shows growing interest in deep learning (DL) and machine learning (ML) models for load prediction [3,4]. The EMS is affected differently depending on the forecast period. Short-term load forecasting (STLF) targets one-day or week-ahead load predictions, enabling reliable power system operation

* Corresponding author.

E-mail address: fukwun@mail.ntust.edu.tw (F.-K. Wang).

through economic dispatch. Maintenance planning and price settling rely on mid-term load forecasting (MTLF). Therefore, a versatile forecasting model can support grid monitoring, optimal scheduling, and energy production and conversion control across various forecast time horizons. DL and hybrid models have demonstrated significant improvements in prediction accuracy in many research areas. Power system load forecasting has been studied extensively over the years. However, emerging technologies like smart grids and energy efficiency strategies still present challenges for load prediction despite extensive research efforts.

The hybrid approach and DL models have gained traction in recent literature and have shown higher performance. They help overcome limitations such as local minimum and extensive training but involve more complex computations [5,6]. The recent integration of renewable energy sources into the grid necessitates more sophisticated models to better understand data patterns. Our proposed model excels at monitoring various types of energy power sources. The ability of artificial intelligence (AI) and hybrid models to learn non-linearly and their fault tolerance position them as excellent decision-makers. Published load forecasting methods that utilize hybrid models have achieved promising results [7–9]. The attention mechanism is an approach that has gained popularity in machine learning models for handling data by focusing on relevant information to improve model performance. The attention layer helps in filtering out the feature's information with high weight. Each value is assigned a weight according to its compatibility with the corresponding key, and the output is computed as a weighted sum of these values.

An effective approach to denoising energy data remains a significant research challenge. Many existing decomposition methods struggle with mode redundancy and fail to identify the optimal number of intrinsic mode functions (IMFs). This results in over-decomposition and increased computational complexity. Additionally, most hybrid models combine multiple deep learning components in parallel or cascade without a clear trade-off between model complexity and computational efficiency. This approach limits their scalability and suitability for real-time applications. Moreover, prior studies have rarely explored diverse types of energy data from various sources, with different temporal characteristics and dimensions. This limits the generalizability and practical application of such models in smart grid systems.

Balance of prediction accuracy, computational cost, overfitting, and effective temporal sequence modelling will be crucial to overcoming the shortcomings of the existing research. This study proposes a novel hybrid deep learning model, termed CEEMDAN-BiLSTM-AM-DNN. It integrates complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN), a bidirectional long short-term memory network with an attention mechanism (BiLSTM-AM), and deep neural network (DNN) layers for short-term energy forecasting. To reduce the noise impact of the collected data for better forecasting due to the noisy nature of time series data, we normalize and then decompose using the CEEMDAN method before modelling. We employ Bayesian optimization based on the Gaussian process (BOGP) [10] for hyperparameter tuning. At the same time, Monte Carlo dropout techniques are used to address prediction uncertainty by accounting for inherent noise in the data. We applied a straightforward and efficient method to determine the optimal number of intrinsic mode functions (IMFs), effectively avoiding over-decomposition, irrelevant components, and elevated computational costs. The proposed model is compared with classical, ML, DL, and other hybrid models.

The main contributions of this paper are summarized as follows.

- (1) Using the BOGP algorithm to tune the hyperparameters and integrating DNN hidden layers to BiLSTM-AM helps to capture more information, which improves short-term energy forecasting.
- (2) To enhance computational efficiency, information extraction in sequential training and prediction performance, we introduce

Algorithm 1, which optimizes the selection of IMFs to generate the most relevant decomposed components. Instead of combining multiple deep learning models, we integrate deep learning layers to improve accuracy. This approach effectively excludes less relevant information, reducing training time and preventing overfitting issues, achieving an average reduction in computational time of 65.78 % for PJM and 56.86 % for wind speed data compared to the default CEEMDAN decomposition method with BiLSTM-AM-DNN.

- (3) The analysis utilizes both univariate and multivariate data with varying time horizons (30 min and 1 h) and various levels of step ahead prediction strategy for validation. Our proposed model consistently delivers superior results, outperforming both traditional and hybrid models.

The remainder of this paper is as follows. Section 2 presents related literature. Section 3 describes the proposed method, followed by experimental prediction results and data analysis. Finally, we make conclusions and future work.

2. Related works

Short-term load forecasting and short-term wind speed forecasting (STWF) are fundamental to the management and operation of smart grid networks [11]. Energy consumption forecasts are categorized into four temporal scopes: ultra-short-term, short-term, medium-term, and long-term outlooks [12,13]. As the power grid advances towards 100 % renewable power integration, robust prediction models are essential for developers seeking optimal sites for new farms and system operators planning power system expansions. DL techniques, renowned for their adeptness in handling large datasets, have found widespread application in load and line loss prediction within contemporary power systems [3, 11,14–16]. Integrating DL methods, including hybrid models, has spurred notable advancements across diverse research domains. Successfully integrating renewable energy sources such as solar and wind into modern smart grid systems depends significantly on accurate short-term forecasts. Ye et al. [17] proposed a physics-guided machine learning based on a spatio-temporal model for wind speed forecasting. Recent research has increasingly emphasized hybrid models due to their improved predictive capabilities across various fields. Different model categories can be hybridized to enhance accuracy [14,18–21]. Aksan et al. [22] proposed a variational mode decomposition (VMD), a convolutional neural network, and long short-term memory (VMD–CNN–LSTM) for forecasting load. A primary obstacle in short-term energy data forecasting is the high noise, discretization, and non-stationarity. To address these challenges, many researchers have proposed utilizing advanced signal processing decomposition techniques like empirical mode decomposition (EMD) [23,24], ensemble empirical mode decomposition (EEMD) [3,25], and VMD [14], singular value decomposition (SVD) [26] and the Savitzky-Golay (SG) filter [27] to smooth time series data for better prediction accuracy. A hybrid deep learning approach was applied based on the optimal LSTM architecture for simultaneous load and electricity price prediction [28]. Fekri et al. [29] proposed an adaptive recurrent neural network that learns from data as they arrive for prediction. Zhang et al. [30] used a grey Lotka-Volterra model (GLVM) to analyze and predict energy consumption trends in categories including nuclear electricity, fossil fuels, and renewable energy. Wind speed or wind power prediction is necessary for an effective functional renewable energy platform. Karijadi et al. [15] applied a hybrid method based on CEEMDAN, empirical wavelet transform (EWT), and used LSTM for ultra-short-term wind power forecasting. Determining the optimal number of IMFs when doing decomposition processing is important because it helps avoid redundant modes and improves computational time. A hybrid model that combines EEMD using an algorithm to determine the number of IMFs, SVR, and BiLSTM-AM (EEMD-SVR-BiLSTM-AM) was proposed in prediction

energy data [3]. Jiang et al. [31] developed a multi-step wind speed forecasting model that integrates CGRU with secondary decomposition (SD), enhancing prediction accuracy. Fu et al. [32] proposed a two-layer decomposition approach for wind speed prediction.

In the context of energy prediction (wind power, wind speed, solar, electricity consumption, etc.), in recent years, transformer models have gained much attention and have shown great results in prediction performance. The transformer model distinguishes itself from earlier architectures by forgoing recurrence and convolution, allowing it to efficiently capture long-range dependencies in data. Chen et al. [33] enhanced the predictive performance of multiple wind farm datasets by integrating a dual attention network into a temporal convolutional network sparse transformer (TCN-DANet-Sparse Transformer), effectively addressing the spatiotemporal coupling of multiple wind farms. To improve the transformer performance, Bentsen et al. [34] developed a fast Fourier transformer for wind speed prediction. An efficient transformer-based LSTF model, Informer, was designed for time series long sequence time-series forecasting [35]. Li et al. [36] incorporated local convolution into transformers and introduced LogSparse attention. Lim et al. [37] introduced the temporal fusion transformer (TFT), an interpretable model for time series analysis that effectively highlights the significance of different variables across various time lags. Li et al. [38] used improved TFT for hourly load prediction. Wu et al. [39] proposed a hybrid method based on VMD, an adaptive differential evolution for hyperparameters tuning of TFT to predict eight real-world 1-h wind speed data sets. Unlike conventional models that merge data from all locations, recent studies have improved spatiotemporal modeling for wind forecasting. For instance, Zhao et al. [40] proposed a location-centric transformer that integrates spatial and temporal

dependencies while preserving site-specific characteristics through spatiotemporal gated fusion and reversible normalization mechanisms for leveraging spectral information. As demonstrated by Sun et al. [41], integrating spatio-temporal correlation analysis with transformer architectures enhances the accuracy of short-term multi-step wind power forecasts with average mean absolute error of 0.0914 and 0.0911 for the two experiments. Zeng et al. [42] proposed an interpretable two-stage decomposition framework for wind speed forecasting, which analyzes component-wise contributions to enhance both accuracy and interpretability. A model name CEEMDAN-CNN-LSTM-SA-AE [43] was used in recent study, which leverages signal decomposition and self-attention to enhance both accuracy and interpretability in household load forecasting.

While many hybrid approaches combine multiple models for prediction and demonstrate strong performance and robustness, most rely on univariate energy data or fail to apply a unique decomposition method to eliminate less relevant components. In contrast, our approach integrates deep learning layers to enhance information gain, improving prediction accuracy and computational efficiency.

3. Methodology

In this section, we discuss the methodological approach of the proposed model. A novel deep learning framework based on CEEMDAN and BiLSTM-AM integrated with a deep neural network at the end of the attention layer is proposed to improve fitting ability and prediction accuracy. Fig. 1 shows the overall analysis for short-term energy forecasting. The model hyperparameters are tuned using the Bayesian optimization method.

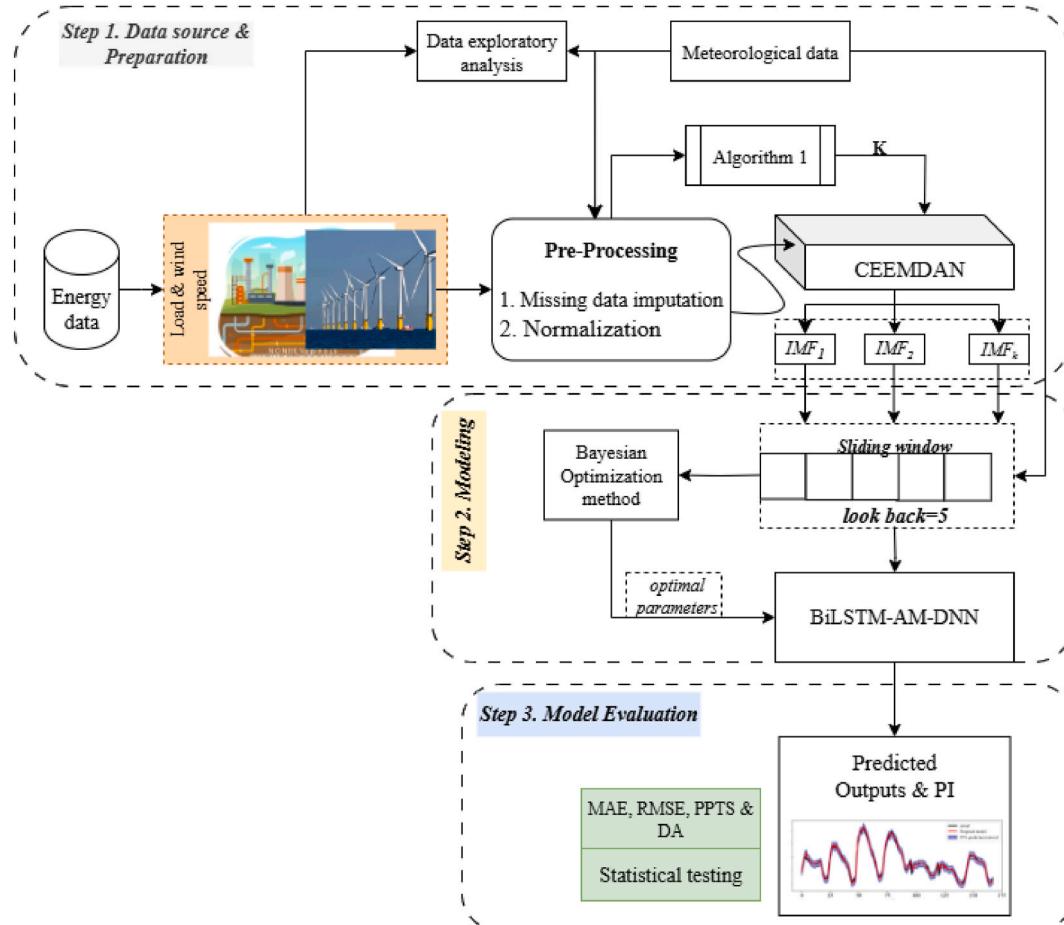


Fig. 1. Framework of the proposed method for energy prediction.

3.1. Complete ensemble empirical mode decomposition with adaptive noise

As a result of new concerns of ensemble EEMD in signal decomposition, such as the effect of residual noise affecting the accuracy of the signal sequence from original data, Torres et al. [44] developed a CEEMDAN algorithm to counter the problem by adding finite adaptive white noise into the original time series during all decomposition processes to overcome the problem of EEMD incompleteness and reconstruction error problem. This method is based on the EMD algorithm [45], which distributes the signal evenly throughout the band at extreme point intervals, and the following steps are involved in the CEEMDAN algorithm.

1. CEEMDAN adds white noise $\varepsilon_i(t)$ to the original signal $x(t)$ with a standard normal distribution $N(0, 1)$. $Y_i(t) = x(t) + \varepsilon_i(t)$, for $i = 1, 2, \dots, N$, where n is the number of ensembles.
2. EMD decomposes the transformed signal $Y_i(t)$ into a set of IMFs and residual.

$$Y_n(t) = \sum_{m=1}^{M-1} IMF_m^{(i)}(t) r_M^{(i)}(t), \quad (1)$$

where, $M - 1$ is the total number of IMFs produced. $IMF_m^{(i)}$ is the m^{th} IMF and $r_M^{(i)}$ the residual at the i^{th} trial.

3. Steps (1) and (2) are repeated for N trials. Each trial adds a different white noise series $\varepsilon_i(t)$ to the original signal.
4. EEMD calculates the last IMF by averaging the total m IMFs across N trials $IMF_m^{\text{ave}}(t) = \frac{1}{N} \sum_{i=1}^N IMF_m^{(i)}(t)$, therefore $IMF_1 = \frac{1}{N} \sum_{i=1}^N IMF_1^{(i)}(t)$ and residual $r_1(t) = x(t) - IMF_1$

Algorithm 1 is applied to determine the number of IMFs. The iteration process applies the coefficient of determination R^2 between the original data training dataset X and mode components (MC_i) and MC_{i+1} by Equation (2). In this study, the energy signal (load or wind speed) is iteratively decomposed into intrinsic mode functions (IMFs) until a predefined stopping criterion is met. The coefficient of determination (R^2) is then calculated for each reconstructed signal, and the optimal number of IMFs is identified as the one corresponding to the lowest R^2 value.

$$R^2 = 1 - \frac{\sum_{t=1}^N (y_t - \hat{y}_t)^2}{\sum_{t=1}^N (y_t - \bar{y}_t)^2} \quad (2)$$

Algorithm 1: Optimum number of IMF selection

Input: Dataset X , m (# of modes by default without residual)
1: $F_x(i) \leftarrow \text{IMFs}_x$
2: Assign: Number of modes $n = 1, \dots, m$ by default setting, where m is the maximum number of IMFs.
3: Decompose X to obtain the maximum IMFs number then delete the EEMD algorithm.
4: for $i = 1$ to m
5: Decompose using mode $_i = \text{CEEMDAN}(X, \text{max_IMF} = i)$
6: Sum mode $_i$ and calculate R^2 between X and mode $_i$
7: if $R_i^2 \leq R_{i+1}^2$ then, repeat Steps 5 and 6.
8: else
9: Break the decomposition process and return i as the optimal number of IMF
10: end if
11: end for
Output: Optimum the number of IMFs

3.2. Bidirectional long short-term memory with attention

The BiLSTM-AM model is used to perform load prediction. BiLSTM is

one of the extended models of LSTM. LSTM solves the long-term dependence on data that RNN cannot effectively train. The memory unit contains self-connection, stores the time state of the network, and is controlled by three gates named “input gate,” “output gate,” and “forgotten gate” [46]. BiLSTM consists of two independent LSTM hidden layers with similar outputs in opposite directions. The previous and future information is contained in the output layer of this architecture. BiLSTM with attention is one of those variants of recurrent neural networks (RNNs). Some good works done in energy market prediction using LSTM and BiLSTM-AM can be found in papers. The formulae in the LSTM are given by

$$f_t = \text{sigmoid}(W_f[M_t, h_{t-1}] + b_f) \quad (3)$$

$$i_t = \text{sigmoid}(W_i[M_t, h_{t-1}] + b_i) \quad (4)$$

$$\tilde{C}_t = \tan h(W_C[M_t, h_{t-1}] + b_C) \quad (5)$$

$$C_t = C_{t-1} \otimes f_t + i_t \otimes \tilde{C}_t \quad (6)$$

$$o_t = \text{sigmoid}(W_o[M_t, h_{t-1}] + b_o) \quad (7)$$

$$h_t = o_t \otimes \tan h(C_t) \quad (8)$$

In Eqs. (3)–(8), $t - 1$ and t are the previous and current time steps. In Eqs. (3), (4) and (7), sigmoid is selected as the activation function, and M_t is the input of LSTM. W_i , W_o , W_f , b_i , b_o , and b_f represent the weights and bias of three gates, respectively. h_{t-1} is the previous hidden unit. After Eqs. (5) and (6) are processed, C_t becomes the current memory unit. Eq. (8) shows the element-wise multiplication of the previous hidden cell output and the previous memory cell. BiLSTM can let input flow in both directions. A forward LSTM and a backward LSTM in BiLSTM share the same input data. A LSTM computes the forward hidden sequence \vec{h} , the backward hidden sequence \overleftarrow{h} and the output sequence y_t by iterating the backward layer and then updating the output layer. The attention mechanism has lately proved effective in various tasks ranging from question answering to machine translation, speech recognition, time series and image captioning [47]. The attention layer helps in filtering out the feature information with high weight. Each value is assigned a weight according to its compatibility with the corresponding key, and the output is computed as a weighted sum of these values. The equations of an attention mechanism are given by

$$\vec{A} = \tan h(W \vec{H} + b) \quad (9)$$

$$\vec{B} = \text{softmax}(\vec{A}) \quad (10)$$

$$F = \sum (\vec{H} * \vec{B}) \quad (11)$$

\vec{H} is the input feature, and W and b represent weight and bias, respectively. In Eq. (10), the softmax function is used to obtain the normalized weight \vec{A} of each value. Finally, the weighted sum of the forward input feature based on the weight \vec{B} is calculated as the forward value F . The F is the output of the attention mechanism. BiLSTM-AM algorithm is further improved by adding a hidden layer of DNN at the end to further process complex data and improve forecasting performance.

3.3. Deep neural network

A deep neural network, also known as a multilayer perceptron, is an artificial neural network composed of multiple layers of interconnected nodes (or “neurons”) situated between the input and output layers [48, 49]. Neurons in each layer are networked and use activation functions and weighted connections to process incoming data. The network learns

more abstract features as input moves through successive layers, which helps it to represent intricate patterns and relationships in the data. Because DNNs can capture complex patterns and generate precise predictions, they are widely employed in many applications, such as natural language processing, predictive analytics, and picture and speech recognition. Fully connected DNN add after attention layer of BiLSTM. The number of neurons for each hidden layer is manually done. The prediction interval is estimated using the Monte Carlo (MC) dropout technique [50].

In this paper we applied a 95 % interval using a MC dropout method η_1^2 and η_2^2 from inference noise approach.

$$\eta_1^2 \leftarrow \frac{1}{N} \sum_{i=1}^N (\hat{y}_{(i)}^* - \bar{y}^*)^2 \quad (12)$$

$$\eta_2^2 \leftarrow \frac{1}{M} \sum_{j=1}^M (\hat{y}_j^* - \bar{y}_v)^2 \quad (13)$$

The total uncertainty estimations are obtained as $\eta \leftarrow \sqrt{\eta_1^2 + \eta_2^2}$ and 95 % prediction interval is obtained as $\text{pred}_{\text{value}} \pm 1.96 \times \eta$. Fig. 2 depicts the architecture of BiLSTM-AM-DNN.

The architecture of Fig. 2 demonstrates the sequential processing of temporal information through the model's components. The raw input data are first processed by the BiLSTM layer, which captures temporal dependencies and sequential patterns. The resulting hidden states are then passed to the attention mechanism, which computes a weighted sum of these states by assigning higher weights to more relevant time steps, thereby emphasizing important temporal features. This attention-refined representation is subsequently fed into the fully connected DNN layers, which transform the compressed information into higher-level feature representations. This final stage enhances the model's representational capacity and enables it to learn complex nonlinear relationships within the data.

4. Dataset and analysis of results

4.1. Dataset

The proposed method is tested on two load datasets with different time intervals and multivariate wind speed data to verify its robustness and generalization. The hourly dataset is from PJM [51], the second data is half-hourly load consumption of New South Wales (NSW) from the Australian Energy Market Operator (AEMO) [52], and an offshore wind

speed using meteorological variables as multivariate data [53]. The wind speed dataset used in this study was originally recorded at 10-min intervals and was converted to hourly intervals. Wind direction, relative humidity, dew point, air temperature, and air pressure were input variables for wind speed prediction. Out of the 25 available stations, VALHALL_A was selected for the multivariate prediction approach in this paper. Fig. 4 (left) highlights 13 stations, with the selected station represented by a green point. The VALHALL_A station data had a few missing observations, imputed using multiple imputations by chained equations [54]. The PJM, NSW, and VALHALL_A datasets shown in Fig. 3 cover the periods from January 2016 to May 2018, January 2018 to August 2021, and January 2023 to March 2024, respectively. The historical wind speed data in Fig. 3 (c) demonstrates increased volatility and randomness. Two types of datasets (i.e., multivariate wind speed as renewable energy and univariate electricity load as non-renewable) are used to validate the performance of the proposed method to forecast for two and seven days. The processed data is maintained in [0, 1] by normalizing the input data. Then, we can solve the problem of inconsistency in training speeds and predictability and speed up the Bayesian optimization algorithm convergence. The normalization method is shown as;

$$X_n = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \quad (14)$$

We analyzed wind direction with wind speed for the selected station, as shown in Fig. 4 (right side). The wind rose diagram reveals prevailing winds from the south to the northwest, indicating a predominant wind direction toward the southwest, west, and northwest. The dominant wind direction is southwest. Descriptive statistics for electricity loads and wind speed are shown in Fig. 5 (a) with positive skewness, showing that most data points are clustered around the left tail. The load and wind speed kurtosis is less than 3, indicating that the distribution is platykurtic. The correlation heatmap in Fig. 5 (b) indicates that only wind direction and relative humidity positively correlate with wind speed, while other factors show a negative relationship.

Given that wind speed is influenced by multiple meteorological factors, both directly and indirectly, we applied a multivariate approach to predict wind speed using these factors as inputs. Since many electricity or renewable energy datasets are not normally distributed, the Pearson correlation coefficient alone may not fully capture the relationships. Therefore, we also used the maximum information coefficient (MIC) to explore the variable relevance. The MIC values for wind

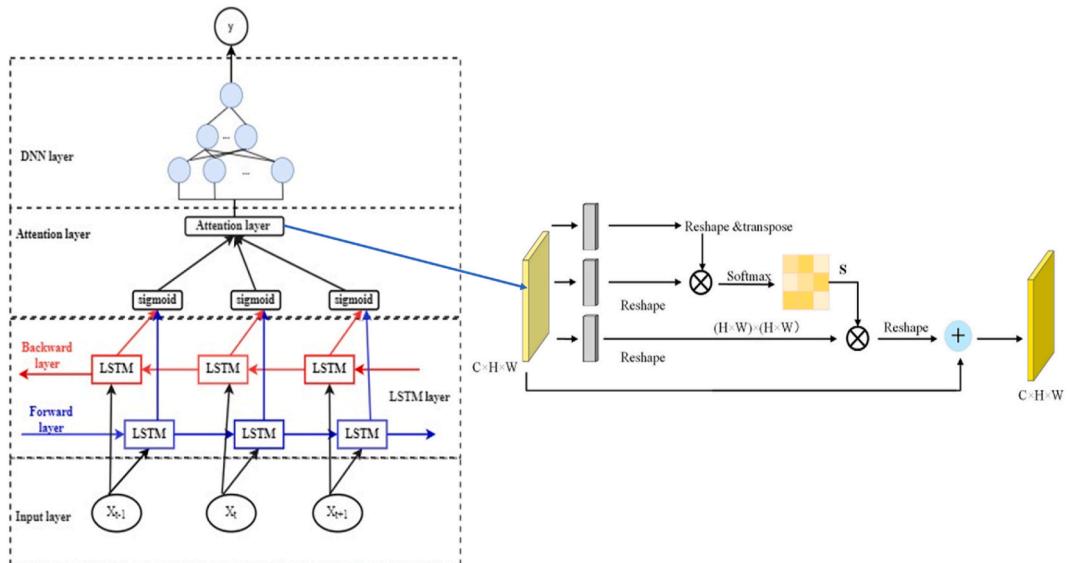


Fig. 2. The architecture of BiLSTM-AM-DNN.

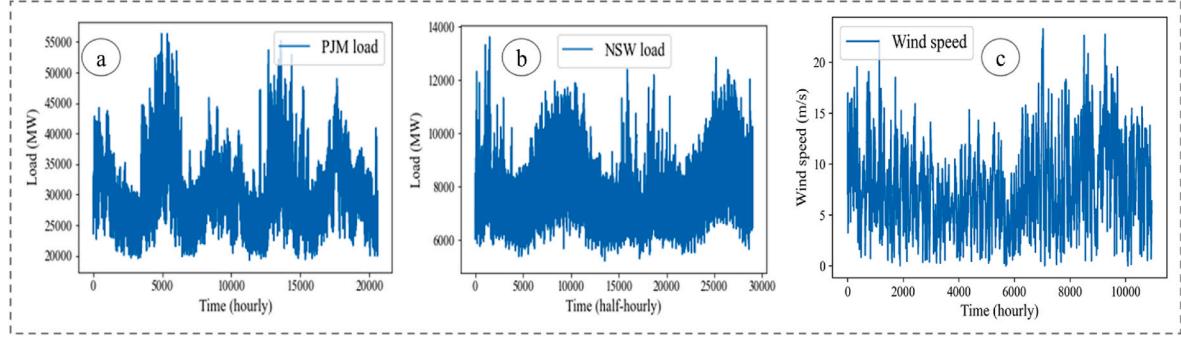


Fig. 3. (a)–(c) PJM, NSW load, and VALHALL_A wind speed datasets.

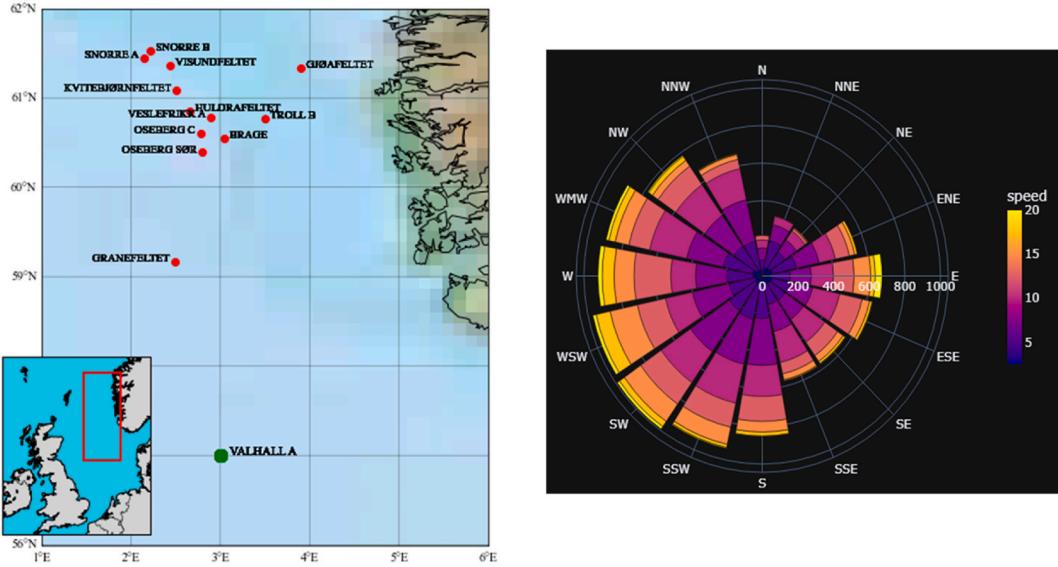


Fig. 4. Map of stations in the North Sea (left), and a rose chart of wind direction for VALHALL A (right).

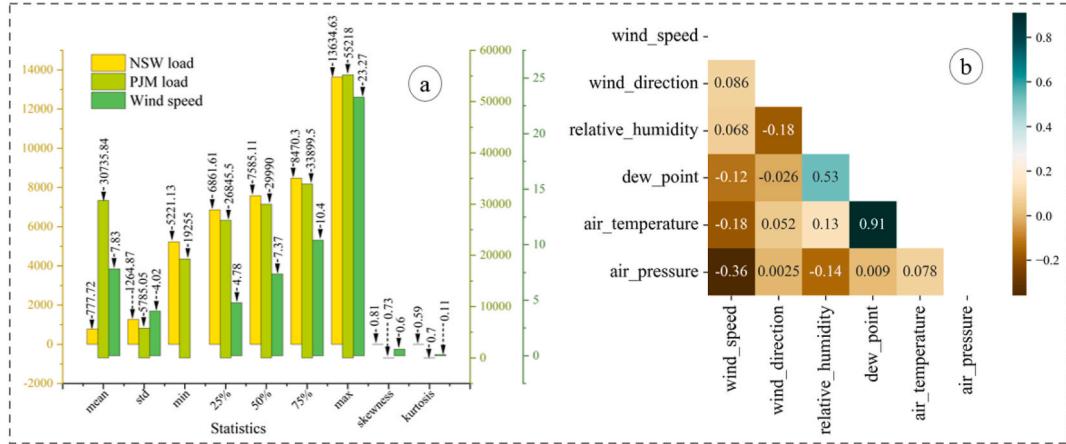


Fig. 5. (a) Descriptive statistics and (b) correlation matrix of VALHALL_A station data.

direction, relative humidity, dew point, air temperature, and air pressure concerning wind speed are 0.0796, 0.0466, 0.0763, 0.0819, and 0.1269, respectively, with air pressure showing the strongest relationship with wind speed. Fig. 6 provides a visual representation of electricity load patterns over different days of the week and hours of the day or across months of the year. In this figure, the x-axis shows the hours of the day (0–23) or months of the year (1–12), while the y-axis indicates

the days of the week (Monday–Sunday). The color intensity reflects the magnitude of electricity consumption. PJM and NSW have higher electricity load demand during the summer months, from June to August. For PJM, daily demand peaks between 1 a.m. and 1 p.m. on weekdays, while for NSW, peak demand occurs between 5 p.m. and 8 p.m. across all days. The PJM and NSW load datasets exhibit pronounced seasonal and daily patterns, coupled with abrupt demand surges that complicate the

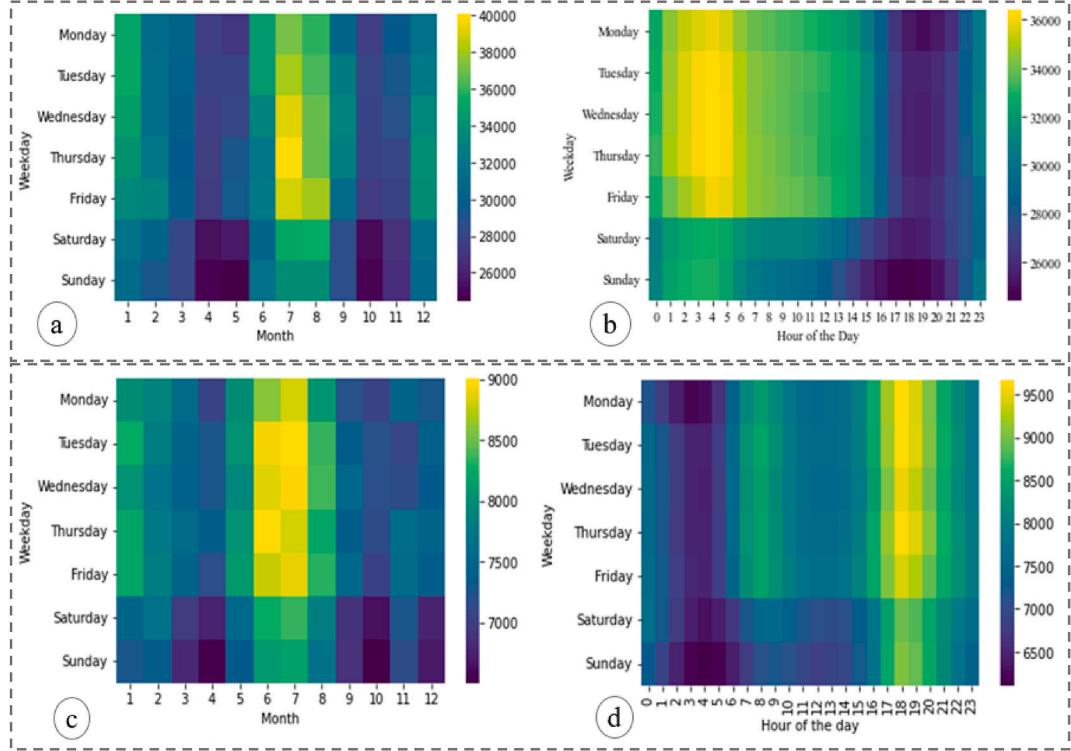


Fig. 6. Heatmap of electricity load by day of the week and month of the year or hour of the day. (a) PJM load day of the week and month of the year; (b) PJM load day of the week and hour of the day; (c) NSW load day of the week and month of the year; (d) NSW load day of the week and hour of the day.

accurate prediction of peak loads. In contrast, the VALHALL_A wind speed dataset is subject to multiple meteorological influences and displays greater stochastic variability, making it inherently more challenging to model, particularly in multivariate forecasting settings.

4.2. Model evaluation indexes and hyperparameter tuning

The implementation of data analysis is in python. TensorFlow is installed in the Spyder 3.9 environment.

Mean absolute error (MAE), directional symmetry or accuracy (DS/DA), peak percentage of threshold statistics (PPTS), PPTS describes the ability to predict peak flows [55] are used to evaluate the prediction effect and improvement rate (IR), which provides a clearer basis for evaluating the relative performance of two models using MAE. Generally, the lower the PPTS, the better the capability to forecast peak flows. To compute the PPTS, the records are organized in descending order. The threshold level γ specifies the percentage of the deleted bottom data from this order, and the parameter G is the number of top data at the threshold level. For example, the PPTS (10 %) criterion evaluates the top 10 % of the peak flows. MAE and RMSE are primarily used to evaluate the accuracy of deterministic forecasts. In contrast, PPTS and DS are primarily used to evaluate the accuracy of peak and directional trend forecasts, respectively.

$$MAE = \frac{\sum_{t=1}^N |y_t - \hat{y}_t|}{N} \quad (15)$$

$$PPTS(\gamma) = \frac{1}{100 - \gamma} \frac{1}{N} \sum_{t=1}^G \left| \frac{y_t - \hat{y}_t}{y_t} \right| \cdot 100 \quad (16)$$

$$DS = \frac{100}{N-1} \sum_{t=2}^n d_t, d_t = \begin{cases} 1, & \text{if } (y_t - y_{t-1})(\hat{y}_t - \hat{y}_{t-1}) > 0 \\ 0, & \text{otherwise} \end{cases} \quad (17)$$

$$IR_{MAE} = \frac{MAE_A - MAE_B}{MAE_A} \times 100 \quad (18)$$

where N denotes the sample size and y_t , \bar{y}_t , and \hat{y}_t denote the actual value, average, and forecasted data samples, respectively. While these indexes can be used to rank models, they cannot be used to determine which model is statistically significantly better than the others. Therefore, we used the Giacomini-White (GW) test [55,57] for conditional predictive ability. In recent studies, the Diebold-Mariano (DM) test has been replaced by GW for model superior efficacy predictive ability [56]. GW is regarded as a generalization of the DM test for unconditional predictive ability, which also accounts for parameter estimation uncertainty through conditioning.

All parameters of the models are tuned using BOGP. The loss function employed is a mean square error (MSE) for the convergence curve during optimization. The training process undergoes 100 maximum iterations. The model is configured to automatically save the current optimal hyperparameters when the loss function value on the validation set reaches its minimum. The BOGP continues to exhibit strong global

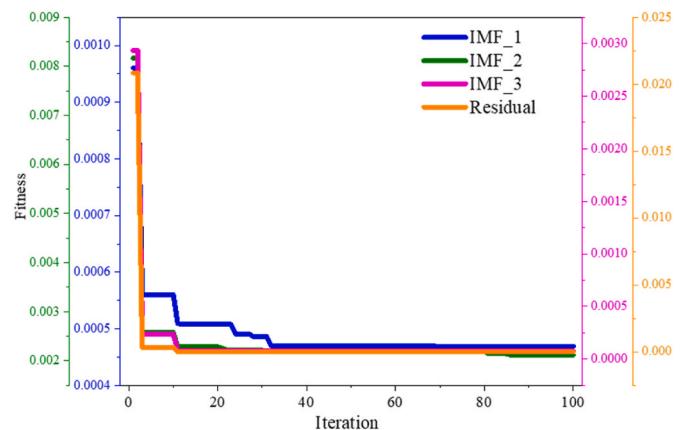


Fig. 7. Convergence curves of BOGP for IMFs and Residual.

exploration capabilities across varying frequency decomposition levels, as demonstrated in Fig. 7. This figure presents the convergence curves for each decomposed mode of the wind data during a 7-day prediction window for single-step forecasting. The results clearly show how BOGP maintains effective exploration despite decomposed modes with distinct frequency characteristics. Table 1 provides the parameter settings of the model using the BOGP approach. Table 2 presents the optimal hyperparameter values for the wind speed and PJM datasets under single-step forecasting horizons of 7-day and 2-day, respectively, while Fig. 8 displays a graphical summary of the single-step wind speed optimal hyperparameter values listed in Table 2, highlighting the top five best-performing trials. The thickest red line with the lowest loss value indicates the iteration corresponding to the best performance with optimal hyperparameter settings. A k-fold cross-validation scheme with $k = 10$ is used during hyperparameter optimization. To balance computational efficiency with high predictive performance, an early stopping strategy was employed to mitigate the risk of overfitting.

Two DNN layers were incorporated after conducting extensive empirical experiments with varying numbers of layers. Stacking two fully connected layers following the attention mechanism enhanced the model's ability to capture high-level nonlinear feature interactions derived from the temporal representations produced by the BiLSTM and attention components. Given the complexity of energy data, a single DNN layer often failed to model these intricate patterns effectively. In contrast, deeper architectures (i.e., more than two layers) resulted in diminishing performance gains and an increased risk of overfitting. The two-layer configuration thus offered an optimal balance between model complexity and generalization performance.

4.3. Results and discussion

As mentioned before, we analyzed three datasets to evaluate the proposed model. Our model is compared with eight benchmark models, which are Autoregressive integrated moving average (ARIMA), SVR, BiLSTM-AM, BiLSTM-AM-DNN, EMD-BiLSTM-AM-DNN, EEMD-BiLSTM-AM-DNN, a recent published hybrid model called CEEMDAN-CNN-LSTM-SA-AE [43], and a hybrid transformer model called CEEMDAN-TFT to verify its effectiveness and superiority. A multi-step prediction strategy is implemented, which is essential for practical engineering applications, as the accuracy of multi-step forecasting intervals plays a critical role in ensuring the reliability and stability of power systems. Our results demonstrate that the proposed model offers strong generalization capability and robustness in accurately forecasting short-term trends for univariate and multivariate datasets. Various performance evaluation metrics and graphical analyses highlight the model's ability to effectively capture peak flow patterns and predict the direction of value changes more accurately than competing models. This showcases its superiority in handling dynamic fluctuations in the data. To evaluate the effectiveness of Algorithm 1, we compared the performance of the proposed model with that of the default CEEMDAN

Table 2

Optimal hyperparameter values.

	Batch size	BiLSTM hidden size	DNN hidden size-1	DNN hidden size-2	Dropout
Single-step 2-day forecast horizon for PJM					
IMF1	19	280	315	232	0.038
IMF2	41	80	400	397	0.171
IMF3	25	26	350	314	0.055
IMF4	21	272	87	403	0.185
Residual	10	484	230	142	0.0007
Single-Step 7-Day Forecast Horizon for VALHALL_A Wind Speed					
IMF1	48	222	312	200	0.041
IMF2	19	35	33	123	0.131
IMF3	34	86	185	404	0.037
Residual	16	335	126	169	0.0004

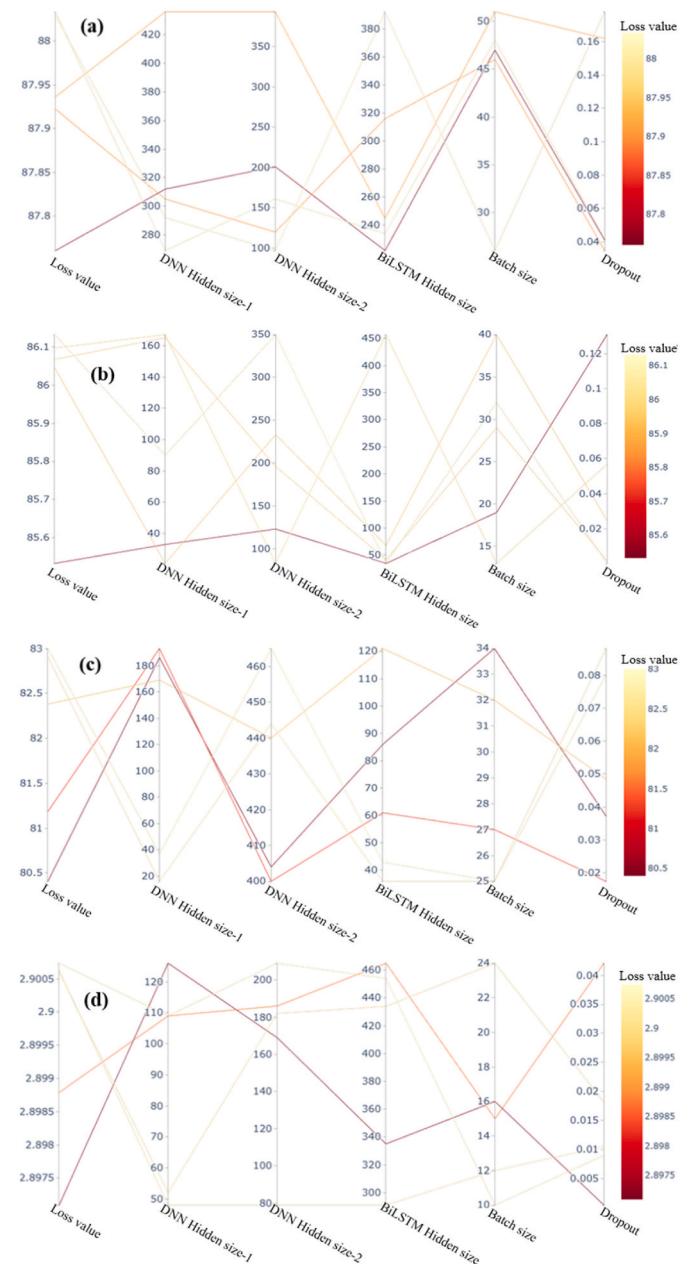


Fig. 8. Parallel coordinate plots during model optimization (top 5 % trials) for Single-Step 7-Day Forecast of Wind Speed: (a) IMF-1; (b) IMF-2; (c) IMF-3; (d) residual.

decomposition method. The default CEEMDAN decomposition method with BiLSTM-AM-DNN is called “Model-1”. Algorithm 1 incorporated to CEEMDAN with BiLSTM-AM-DNN is called “Proposed model”.

This comparison allowed us to assess how well the proposed approach performs relative to a standard decomposition technique. The proposed Algorithm 1 with our model demonstrated significant improvements over the default CEEMDAN method, with performance gains of 38.46 % and 4.60 % for two- and seven-day forecasts of PJM, respectively, and reduced computing time. Similarly, the improvements were even more pronounced for wind speed predictions, achieving enhancements of 69.24 % and 82.35 % for the two- and seven-day forecasts. The evaluation metrics for the models, presented in Tables 3–5, highlight that the proposed CEEMDAN-BiLSTM-AM-DNN model for test data significantly enhances forecasting performance compared to other approaches. In comparison to the CEEMDAN-BiLSTM-AM model, which lacks DNN layers, our proposed model achieves percentage improvements of 22.93 %, 16.27 %, and 4.52 % for the two-day forecast based on 1-step, 5-step, and 10-step ahead predictions, respectively of PJM in Table 3. This demonstrates the substantial advantage of incorporating DNN layers in enhancing predictive accuracy. Figs. 9 and 10 show the prediction results of different prediction steps ahead (1, 5, and 10 steps) of PJM load and VALHALL_A wind speed during the two-day and seven-day prediction period. Across all data types, our model demonstrates superior capability in accurately capturing peak values and predicting trends compared to competing models, as evidenced by the PPTS and DA metrics in Tables 3 and 4. This highlights the model’s effectiveness in both precision and directional accuracy. As shown in Figs. 9 and 10, the prediction step length impacts forecasting performance. As shown in Tables 3 and 4, single-step predictions consistently yield higher accuracy than multi-step forecasts, with a noticeable decline in performance as the prediction horizon increases. This degradation is primarily due to the accumulation of errors over successive prediction steps, a well-known limitation of recursive forecasting approaches [58]. Accumulated errors result from the compounding of small deviations at each prediction step, which can ultimately lead to substantial divergence in long-term forecasts. Nevertheless, the proposed model exhibited superior robustness compared to benchmark models, maintaining relatively

stable mean absolute error, and directional accuracy (DA) across varying forecast lengths. The smaller error differences observed between prediction steps indicate that the hybrid architecture of the proposed model effectively mitigates, though not eliminate, the performance deterioration typically associated with multi-step forecasting. Similar prediction performance is shown in Table 6 for a single-step ahead forecasting of NSW load data.

The comparison between our proposed model and model-1 reveals a significant improvement in both computational efficiency and performance accuracy shown in Table 5. The proposed model effectively reduces less relevant decomposed components, thereby saving time and preventing overfitting issues. This optimization not only accelerates model training but is also crucial for practical engineering applications, where efficiency and reliability are paramount. The percentage improvement in RMSE for the two-day and seven-day forecasts of the proposed model compared to model-1 is 34.17 % and 5.37 % for PJM and 53.01 % and 49.2 % for wind data, respectively. Additionally, computational time efficiency improved by an average of 65.78 % for PJM and 49.23 % for VALHALL_A wind data in case 2 compared to case 1.

The proposed model demonstrated strong predictive performance with an average MAE of 1114.657 for a two-day forecast and 1229.927 for a seven-day forecast horizon, along with DA values of 0.74 and 0.81, respectively, for PJM load data. In comparison, the hybrid models CEEMDAN-BiLSTM-AM and EEMD-BiLSTM-AM-DNN had higher MAEs, with CEEMDAN-BiLSTM-AM recording 1246.263 and 1701.76 and EEMD-BiLSTM-AM-DNN showing 1320.403 and 1461.33 for the same forecasting intervals. This highlights the superior accuracy and efficiency of the proposed model over its counterparts, and better accuracy can be obtained with shorter-term prediction compared to longer periods. As measured by MAE, the proposed models demonstrated significant improvement over hybrid models in DA and forecasting performance. These evaluation metrics highlight the enhanced predictive ability of the proposed approach.

Regarding DA for predictive trends, the proposed model demonstrated a 5.26 % improvement over CEEMDAN-BiLSTM-AM for the seven-day forecast of PJM data. It also showed a 1.27 % improvement

Table 3
Comparison metric results of different models at different prediction steps for PJM data.

Predicted steps	Models	Two days forecast				Seven days forecast			
		MAE	RMSE	PPTS	DA	MAE	RMSE	PPTS	DA
1 step	ARIMA	447.63	968.75	0.0006	0.79	523.83	1081.62	0.0010	0.83
	SVR	476.63	994.45	0.0003	0.81	514.38	1048.66	0.0013	0.83
	BiLSTM-AM	452.54	997.43	0.0004	0.82	513.90	1089.17	0.0014	0.84
	BiLSTM-AM-DNN	439.02	974.54	0.0003	0.85	535.60	1092.07	0.0012	0.84
	EEMD-BiLSTM-AM-DNN	258.31	307.67	0.0005	0.83	251.82	324.79	0.0007	0.94
	CEEMDAN-BiLSTM-AM	307.09	413.30	0.0003	0.93	372.48	531.38	0.0013	0.86
	CEEMDAN-TFT	239.49	375.72	0.0009	0.83	394.61	576.51	0.0019	0.89
	CEEMDAN-CNN-LSTM-SA-AE [43]	297.58	581.42	0.0010	0.83	631.43	940.39	0.0041	0.80
	Proposed model	236.67	340.77	0.0003	0.89	356.45	485.11	0.0012	0.86
5 steps	ARIMA	2151.7	2705.0	0.0020	0.57	2324.19	3207.48	0.0094	0.59
	SVR	2340.1	2868.7	0.0043	0.57	2560.46	3280.28	0.0101	0.60
	BiLSTM-AM	2051.4	2920.1	0.0024	0.66	2229.44	2974.50	0.0099	0.60
	BiLSTM-AM-DNN	1991.3	2796.4	0.0018	0.57	2240.72	3161.98	0.0096	0.64
	EEMD-BiLSTM-AM-DNN	1576.9	1909.9	0.0067	0.60	1817.89	2377.98	0.0054	0.71
	CEEMDAN-BiLSTM-AM	1441.6	1825.3	0.0054	0.68	2012.04	2650.59	0.0056	0.71
	CEEMDAN-TFT	1223.7	1431.9	0.0077	0.68	1608.18	2084.01	0.0037	0.65
	CEEMDAN-CNN-LSTM-SA-AE [43]	1608.7	2062.7	0.0036	0.55	1742.52	2279.56	0.0063	0.66
	Proposed model	1207.1	1505.5	0.0046	0.64	1575.00	2141.15	0.0040	0.76
10 steps	ARIMA	2568.5	3366.9	0.003	0.57	3625.29	4644.74	0.0209	0.54
	SVR	2555.6	3301.3	0.004	0.55	3664.71	4542.41	0.0196	0.55
	BiLSTM-AM	2238.9	2899.6	0.003	0.55	3793.29	4766.68	0.0261	0.55
	BiLSTM-AM-DNN	2274.8	3162.7	0.001	0.60	3644.07	4601.66	0.0207	0.56
	EEMD-BiLSTM-AM-DNN	2126.0	2600.3	0.018	0.68	2314.28	3036.76	0.0070	0.72
	CEEMDAN-BiLSTM-AM	1990.1	2483.2	0.013	0.64	2720.76	3423.40	0.0068	0.71
	CEEMDAN-TFT	1614.3	1933.1	0.005	0.66	2338.84	2859.17	0.0098	0.69
	CEEMDAN-CNN-LSTM-SA-AE [43]	2000.9	2244.5	0.008	0.70	2420.14	3181.54	0.0140	0.66
	Proposed model	1900.2	2143.0	0.011	0.70	1758.33	2200.02	0.0059	0.78

Table 4

Comparison metric results of different models at different prediction steps for VALHALL_A station.

Predicted steps	Models	Two days forecast				Seven days forecast			
		MAE	RMSE	PPTS	DA	MAE	RMSE	PPTS	DA
1 step	ARIMA	0.5243	0.7189	0.0090	0.70	0.5991	0.7668	0.0070	0.65
	SVR	0.5128	0.7126	0.0102	0.66	0.5854	0.7527	0.0072	0.66
	BiLSTM-AM	0.4962	0.6935	0.0081	0.70	0.4799	0.5483	0.0065	0.69
	BiLSTM-AM-DNN	0.4937	0.6989	0.0074	0.66	0.3770	0.5170	0.0063	0.72
	EEMD-BiLSTM-AM-DNN	0.2334	0.2974	0.0049	0.85	0.2492	0.3253	0.0013	0.85
	CEEMDAN-BiLSTM-AM	0.2372	0.3257	0.0082	0.83	0.2531	0.3282	0.0014	0.85
	CEEMDAN-TFT	0.2241	0.3062	0.0050	0.79	0.3902	0.5019	0.0030	0.80
	CEEMDAN-CNN-LSTM-SA-AE [43]	0.2999	0.3962	0.0037	0.72	0.4972	0.6194	0.0056	0.69
5 steps	Proposed model	0.1658	0.2246	0.0028	0.89	0.1726	0.2189	0.0016	0.89
	ARIMA	1.4231	1.7989	0.0044	0.51	1.7050	2.0858	0.0204	0.49
	SVR	1.3489	1.6924	0.0075	0.45	1.6933	2.0762	0.0196	0.50
	BiLSTM-AM	1.4337	1.7392	0.0063	0.53	1.6143	2.0417	0.0191	0.52
	BiLSTM-AM-DNN	1.3982	1.7206	0.0051	0.49	1.5618	2.0237	0.0186	0.50
	EEMD-BiLSTM-AM-DNN	0.4967	0.6328	0.0033	0.70	0.6873	0.8861	0.0014	0.69
	CEEMDAN-BiLSTM-AM	0.6035	0.7589	0.0102	0.66	0.6927	0.8795	0.0015	0.71
	CEEMDAN-TFT	0.4437	0.5721	0.0024	0.64	0.6639	0.8388	0.0067	0.69
10 steps	CEEMDAN-CNN-LSTM-SA-AE [43]	0.7324	0.8993	0.0088	0.68	1.0479	1.3673	0.0120	0.66
	Proposed model	0.6644	0.8079	0.0036	0.72	0.6563	0.8034	0.0019	0.69
	ARIMA	2.4126	2.8071	0.0143	0.60	2.5789	3.1176	0.0400	0.45
	SVR	2.1348	2.5949	0.0165	0.47	2.5753	3.1521	0.0415	0.54
	BiLSTM-AM	2.1117	2.4691	0.0143	0.32	2.5810	3.1051	0.0400	0.55
	BiLSTM-AM-DNN	2.0714	2.4432	0.0147	0.43	2.5485	3.1595	0.0402	0.55
	EEMD-BiLSTM-AM-DNN	1.6502	2.0321	0.0077	0.70	1.6097	1.9887	0.0227	0.57
	CEEMDAN-BiLSTM-AM	1.3738	1.7464	0.0089	0.51	1.3279	1.6431	0.0137	0.68
20 steps	CEEMDAN-TFT	0.8460	1.0073	0.0053	0.66	1.0680	1.2718	0.0113	0.63
	CEEMDAN-CNN-LSTM-SA-AE [43]	0.9019	1.0885	0.0039	0.68	1.0600	1.3484	0.0127	0.63
	Proposed model	0.8843	1.2274	0.0099	0.70	1.2346	1.6004	0.0129	0.64

Table 5

Comparison results of proposed model and model-1.

Dataset	Model	Two days forecast				Seven days forecast			
		MAE	RMSE	PPTS	Time (mins.)	MAE	RMSE	PPTS	Time (mins.)
PJM	Model-1	384.59	517.62	0.001	18.73	373.62	512.62	0.0013	23.36
	Proposed model	236.67	340.77	0.0003	6.20	356.45	485.11	0.0012	7.29
Wind speed	Model-1	0.361	0.478	0.0072	14.41	0.340	0.431	0.0042	11.88
	Proposed model	0.1658	0.2246	0.0028	7.16	0.1726	0.2189	0.0016	6.16

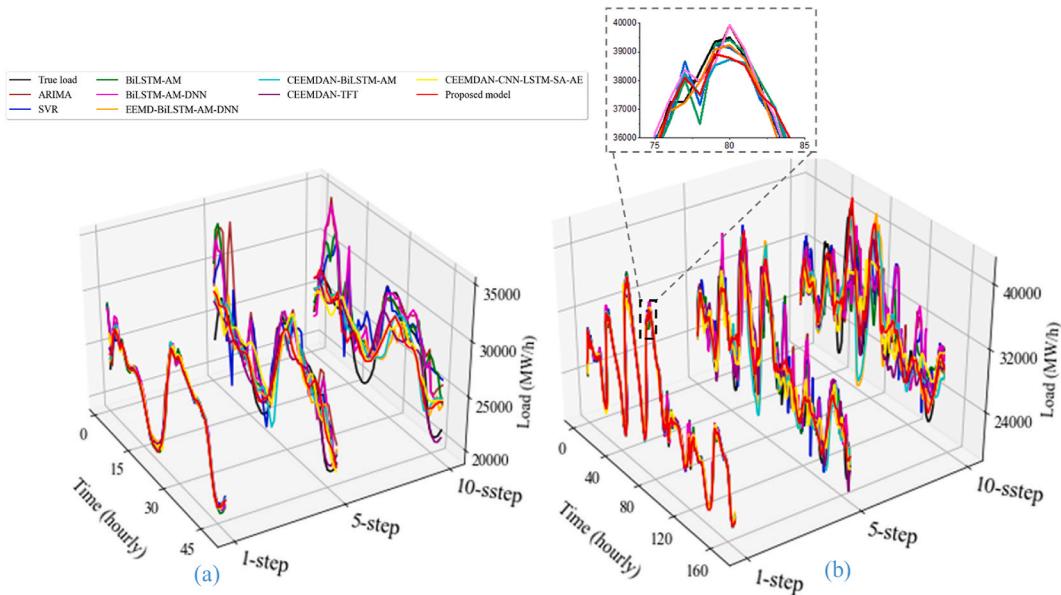


Fig. 9. The PJM load multi-step ahead forecasting results: (a) two-day forecast; (b) seven-day forecast.

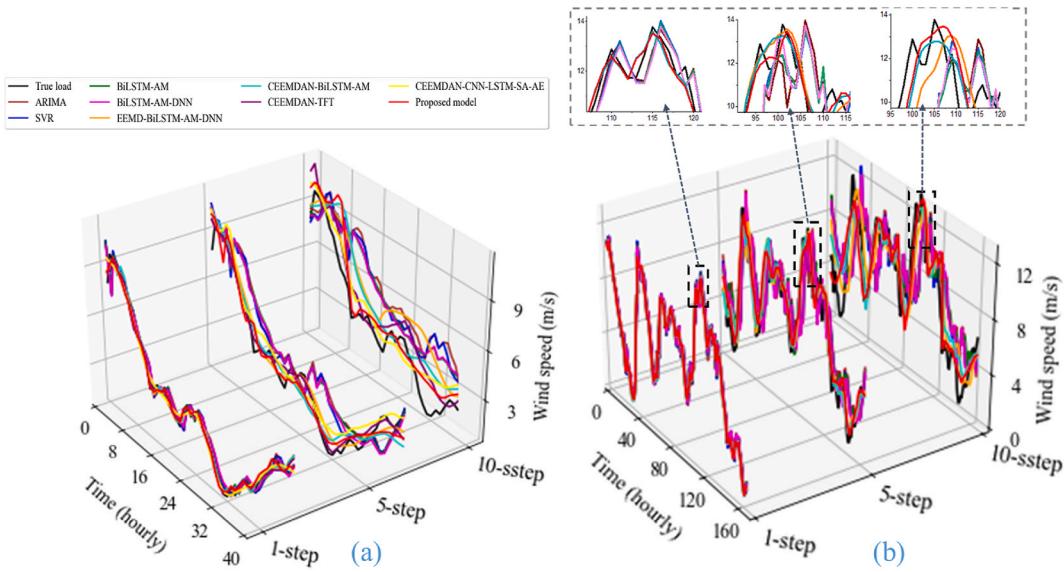


Fig. 10. Multi-step ahead forecasting results of VALHALL_A station: (a) two-day forecast; (b) seven-day forecast.

Table 6

Comparison performance evaluation indexes for NSW load.

Predicted steps	Model	Two days forecast			Seven days forecast		
		MAE	PPTS	DA	MAE	PPTS	DA
1 step	ARIMA	107.28	0.0071	0.85	102.43	0.0009	0.89
	SVR	109.57	0.0013	0.89	94.74	0.0006	0.92
	BiLSTM-AM	91.42	0.0010	0.89	93.42	0.0006	0.93
	BiLSTM-AM-DNN	90.35	0.0007	0.89	88.45	0.0005	0.93
	EEMD-BiLSTM-AM-DNN	52.92	0.0013	0.91	74.25	0.0005	0.93
	CEEMDAN-BiLSTM-AM	37.41	0.0004	0.98	42.25	0.0005	0.93
	Proposed model	47.56	0.0006	0.96	31.58	0.0001	0.95

for the two-day forecast and a 5.71 % improvement for the seven-day forecast compared to EEMD-BiLSTM-AM-DNN. The transformer-based hybrid model CEEMDAN-TFT outperforms the proposed model in both 5- and 10-step-ahead forecasts for two-day forecasts and also in 10-step-ahead for seven-day prediction horizons. However, it falls short in accurately capturing the directional trend of wind speed. In contrast, the proposed model demonstrates superior directional accuracy, as indicated by higher DA values presented in Table 4. A similar scenario can be observed in renewable data (wind speed) evaluation results, underscoring the enhanced forecasting performance of the proposed model in capturing directional trends. Fig. 11 illustrates the reliability of the proposed model in estimating uncertainty through Monte Carlo dropout, presenting the 95 % prediction interval for the seven-day forecast alongside the test data. The model consistently achieves empirical coverage rates exceeding 95 % across all one-step-ahead forecast

horizons for all datasets, demonstrating strong robustness. Notably, the prediction coverage in Fig. 11 reaches 96.58 %, confirming the model's ability to generate reliable probabilistic forecasts. Such prediction intervals offer grid operators critical insights into forecast uncertainty, supporting informed decision-making in operational planning, reserve allocation, and risk-aware energy dispatching. As shown in Fig. 12, the prediction error of MAE for the single models increases gradually as the number of prediction steps ahead rises, in contrast to the proposed model and other hybrid models. Notably, the proposed model is less affected by the increase in prediction steps.

In addition to using performance evaluation metrics such as MAE and RMSE, detailed statistical tests were conducted based on the Giacomini-White (GW) test to benchmark the models. The GW test results, shown as heat maps with GW scores and their p-values, are structured as chessboards using a multivariate GW test with the L1

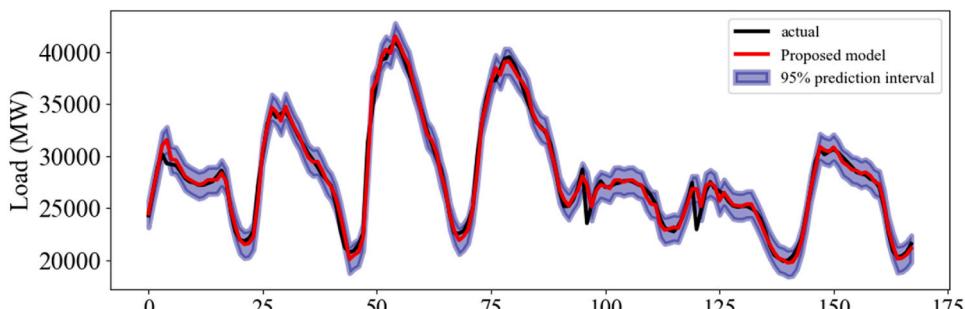


Fig. 11. Comparison plot of the proposed model for seven-day forecast of PJM dataset, with 95 % prediction band in the blue area.

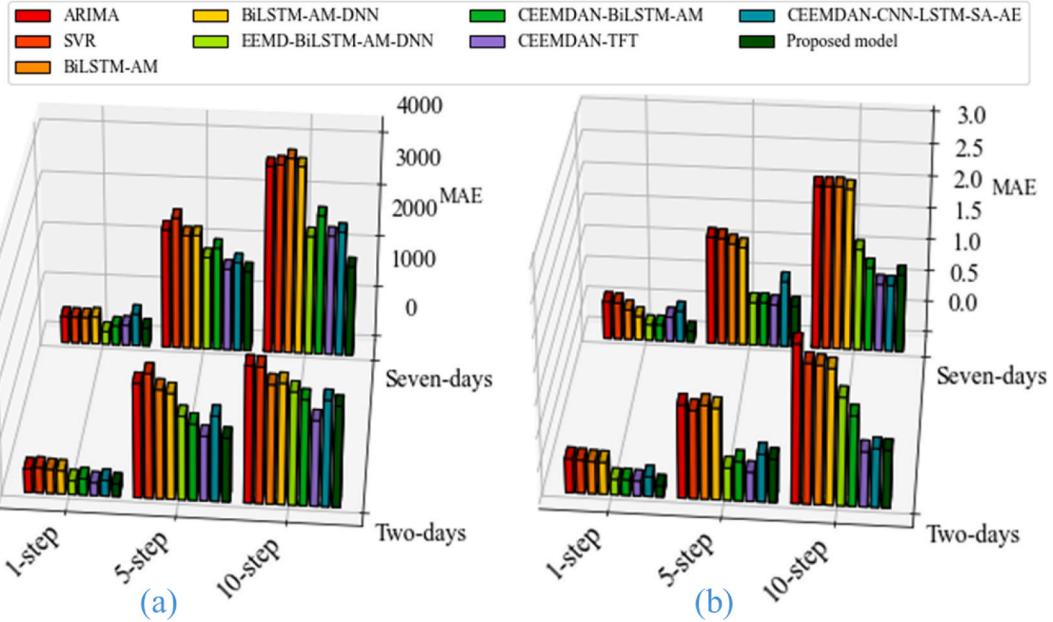


Fig. 12. Prediction error index based on MAE: (a) PJM load data; (b) VALHALL_A station data.

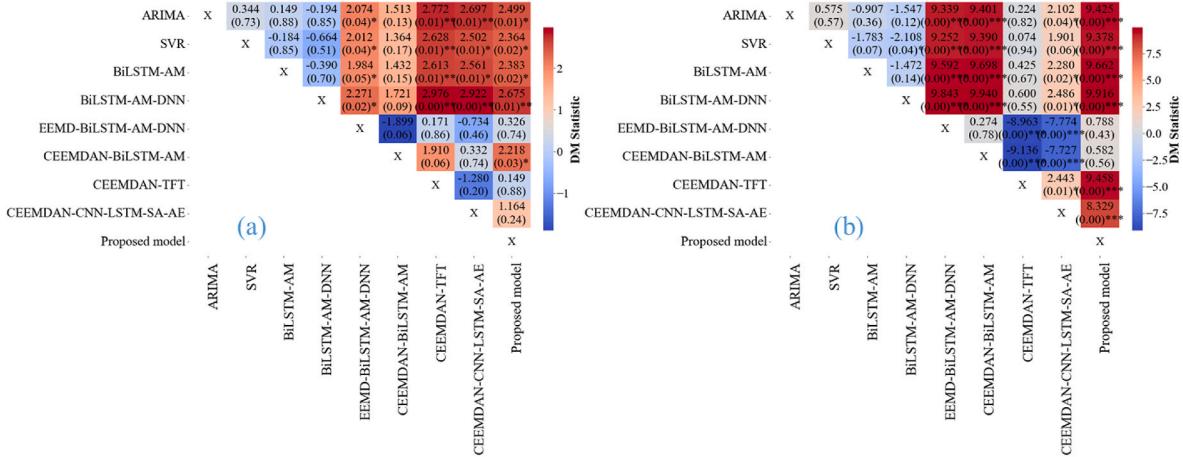


Fig. 13. Heat map GW results for pairwise comparison: (a) PJM single step two-day forecast; (b) Wind data single step seven-day forecast.

norm, as seen in Figs. 13 and 14. In Fig. 13(a) and (b), the GW test results at a significance level of 0.05 indicate that the proposed model statistically outperforms other models in many cases, with p-values of 0.001***, 0.01**, and 0.05*. A higher GW score paired with a lower p-value suggests superior model performance in the pairwise comparison, where the X-axis represents the superior model and the Y-axis the inferior model. White squares indicate inconclusiveness between models, as seen in Fig. 13(a) between the proposed model and EEMD-BiLSTM-AM-DNN. Although all hybrid models performed well on average in Fig. 12(b), the proposed model remained the best, particularly during the seven-day forecast for VALHALL_A station's renewable energy data. Fig. 13 displays the prediction results for PJM over the seven-day forecast horizon (Fig. 14(a)–(c)) and for VALHALL_A station over the two-day forecast period (Fig. 14(d)–(f)) for all prediction steps. In Fig. 14 darker green square indicate a lower p-value closer to zero indicating more significant difference between the forecasts of a model on the X-axis (better) and the forecast of a model on the Y-axis (worse). In Fig. 14(f), for the 10-step ahead prediction of renewable wind data over the two-day horizon, the first row is entirely green, indicating that the ARIMA model is significantly outperformed by all other models.

In Figs. 13 and 14, the last column is predominantly red or green, indicating that the proposed model performs statistically significantly better than the competing models. A few exceptions are observed in Fig. 13(a), where the difference between the proposed model and EEMD-BiLSTM-AM-DNN is not statistically significant, and in Fig. 14(d), where the same applies to CEEMDAN-BiLSTM-AM. Additionally, Fig. 14(a) and (e) show that EEMD-BiLSTM-AM-DNN outperformed the proposed model in those specific cases.

The ablation analysis, presented in Figs. 15 and 16, evaluates the predictive performance of the proposed model and its sub-models based on loss metrics. This analysis investigates the individual contributions of the CEEMDAN decomposition algorithm, the attention mechanism, and the DNN layers to the overall forecasting accuracy. While the full CEEMDAN-BiLSTM-AM-DNN model consistently demonstrates superior performance across scenarios, the DNN layers, in many cases, further enhance the temporally processed features more effectively than the attention mechanism alone. Fig. 16 illustrates the percentage improvement of the proposed model over its sub-models, with notable gains observed particularly in comparisons with BiLSTM-AM and BiLSTM-DNN, underscoring the critical role of data decomposition.

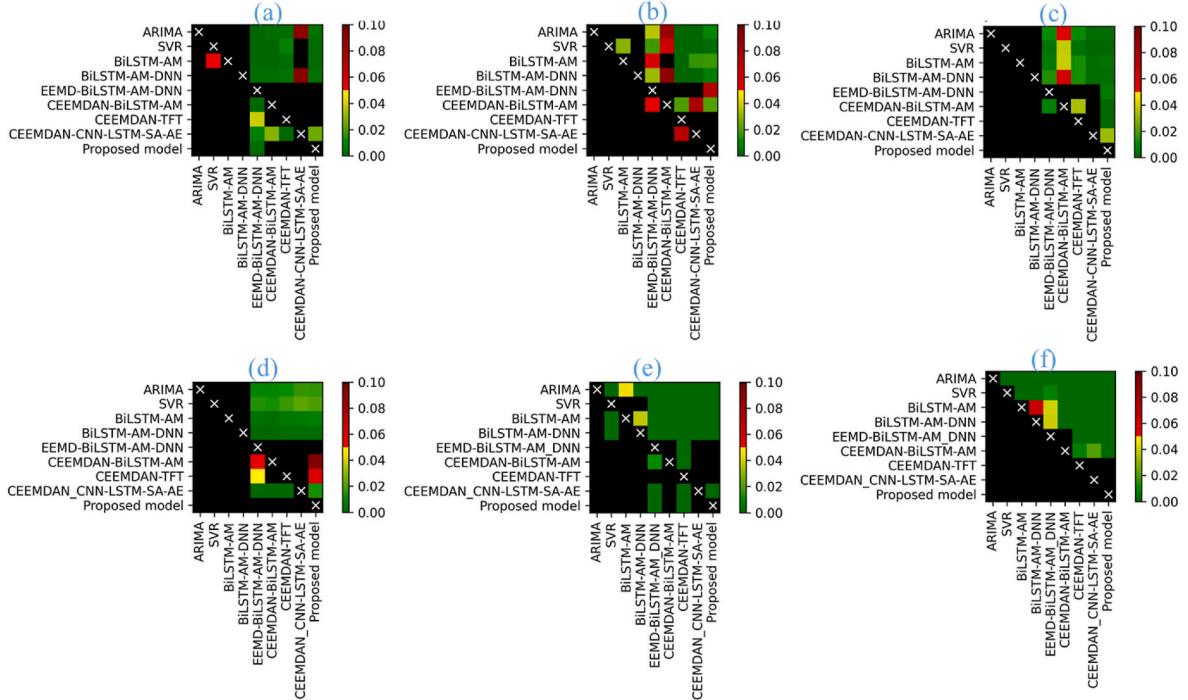


Fig. 14. Results of GW test with multivariate loss: (a)–(c) represents PJM seven-day forecast for all steps ahead, and (d)–(f) represents a two-day forecast for VALHALL_A station wind data. Darker green squares denote a p-value closer to zero, indicating a more statistically significant difference between the model on the X-axis (better) and the model on the Y-axis (worse).

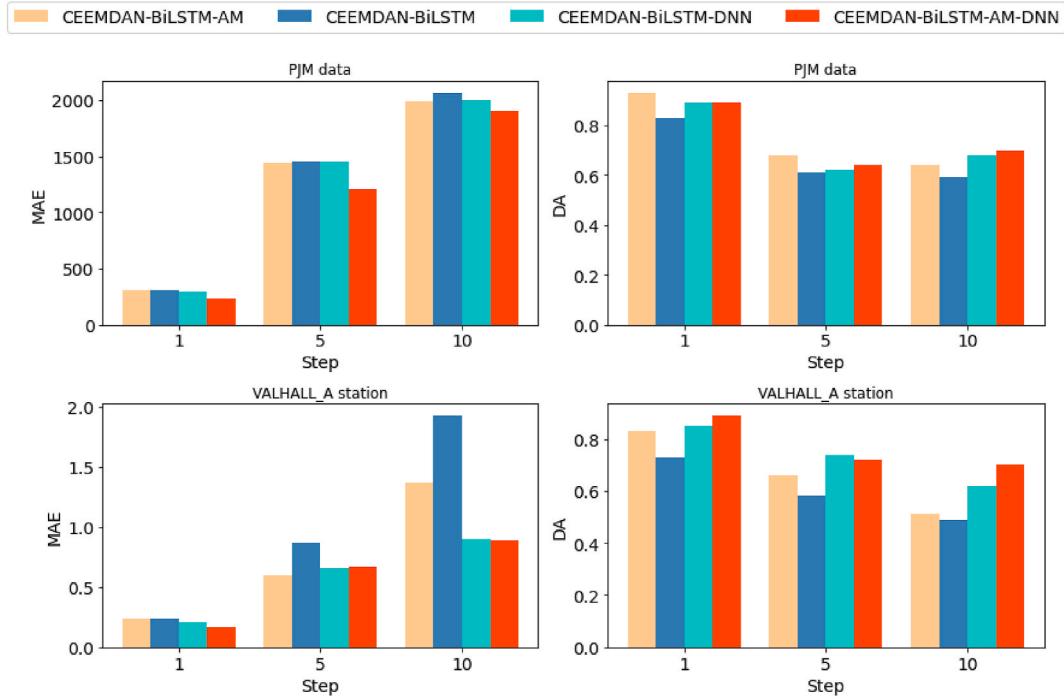


Fig. 15. Comparison of the evaluation score of sub-models and proposed model during two-day forecast.

Specifically, the CEEMDAN method improves prediction performance by up to twofold compared to models without decomposition. The CEEMDAN-BiLSTM model exhibits a substantially higher MAE at the 10-step horizon, indicating limited scalability for long-term forecasting. In contrast, the proposed model achieves superior DA, effectively capturing trend directions in both load demand and wind speed, which is essential for informed operational decision-making. These findings

affirm that each module provides substantial benefits, emphasizing the effectiveness of integrating complementary components within a hybrid modeling framework.

5. Conclusions

The development of smart grids and corporate renewable energy

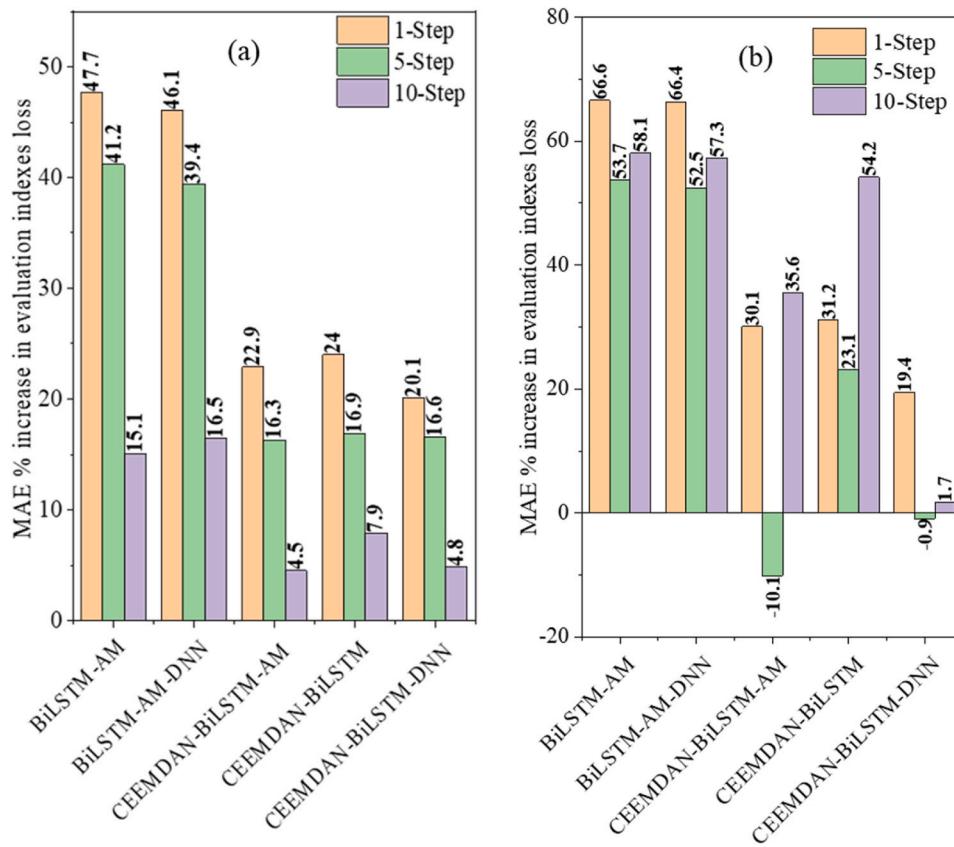


Fig. 16. Changes in metrics losses between the proposed model and sub-models across ablation tests: (a) PJM two-day forecast; (b) Wind speed two-day forecast.

sources has significantly increased in recent years. With growing efforts to reduce emissions and optimize resource use, integrated energy systems have become crucial. For effective smart grid operation, robust and generalizable models for forecasting multiple energy sources are essential. This study proposes a novel hybrid model combining an optimal IMF-determining decomposition method, a BiLSTM-AM architecture with two DNN layers, and a BOGP approach for hyperparameter tuning. The proposed approach is explored and tested on three multi-energy datasets with different time intervals to see its reliability, generalization, and robustness in handling data of different frequency flow patterns. MC dropout method is applied for the prediction interval to validate the reliability of the proposed model. An uncertainty estimation method at scale was presented, covering 95 % of an uncertain forecast. The analysis results on the test set show that the proposed algorithm achieves the highest prediction accuracy, with an average PPTS value below 0.005 across all prediction periods. Additionally, it attains an average DA of over 75 % for VALHALL_A station data and above 80 % for the PJM load forecast with an empirical prediction coverage of 95 %. Our model's computational efficiency and speed based on the IMFs determining algorithm compared to a default decomposition approach significantly improved by an average of 65.78 % for PJM and 56.86 % for VALHALL_A renewable wind data. A sliding window approach is used during forecasting. Experimental results show that: 1) Determining the optimal number of decomposed components reduces training time and enhances prediction accuracy. The model avoids overfitting issues by excluding less relevant components from including unnecessary data. 2) Incorporating DNN layers into the BiLSTM-AM model enhances its ability to capture more relevant features, resulting in improved prediction performance. This study highlights the importance of integrating meteorological factors with advanced machine learning models to accurately predict smart grid energy demand.

Although the proposed model demonstrates strong performance

through the integration of an optimal IMF selection algorithm during decomposition and the incorporation of two DNN layers into the BiLSTM-AM architecture, the study lacks an in-depth comparison with more advanced decomposition techniques, such as Improved CEEMDAN and VMD. Future research will focus on further optimizing the architecture by determining the ideal number of DNN layers, ensuring robust performance across varying forecast horizons, and benchmarking the model against more sophisticated decomposition methods to enhance the reliability and efficiency of energy demand forecasting in smart grid applications.

CRediT authorship contribution statement

William Gomez: Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Data curation. **Fu-Kwun Wang:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Conceptualization. **Shey-Huei Sheu:** Writing – review & editing, Supervision.

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 will be made available on request.

References

- [1] Bodong S, Wiseong J, Chengmeng L, Khakichi A. RETRACTED:Economic management and planning based on a probabilistic model in a multi-energy market

- in the presence of renewable energy sources with a demand-side management program. *Energy* 2023;269:126549. <https://doi.org/10.1016/j.energy.2022.126549>.
- [2] Li C, Li G, Wang K, Han B. A multi-energy load forecasting method based on parallel architecture CNN-GRU and transfer learning for data deficient integrated energy systems. *Energy* 2022;259:124967. <https://doi.org/10.1016/j.energy.2022.124967>.
- [3] Gomez W, Wang F-K, Lo S-C. A hybrid approach based machine learning models in electricity markets. *Energy* 2024;289:129988. <https://doi.org/10.1016/j.energy.2023.129988>.
- [4] Fan G-F, Zhang L-Z, Yu M, Hong W-C, Dong S-Q. Applications of random forest in multivariable response surface for short-term load forecasting. *Int J Electr Power Energy Syst* 2022;139:108073. <https://doi.org/10.1016/j.ijepes.2022.108073>.
- [5] Khan PW, Byun Y-C, Lee S-J, Kang D-H, Kang J-Y, Park H-S. Machine learning-based approach to predict energy consumption of renewable and nonrenewable power sources. *Energy* 2020;13:4870. <https://doi.org/10.3390/en13184870>.
- [6] Zhang J, Wei Y-M, Li D, Tan Z, Zhou J. Short term electricity load forecasting using a hybrid model. *Energy* 2018;158:774–81. <https://doi.org/10.1016/j.energy.2018.06.012>.
- [7] Aslam S, Herodotou H, Mohsin SM, Javaid N, Ashraf N, Aslam S. A survey on deep learning methods for power load and renewable energy forecasting in smart microgrids. *Renew Sustain Energy Rev* 2021;144:110992. <https://doi.org/10.1016/j.rser.2021.110992>.
- [8] Zhou M, Zhu Z, Hu F, Bian K, Lai W, Hu T. Short-term commercial load forecasting based on peak-valley features with the TSA-ELM model. *Energy Sci Eng* 2022;10:2622–36. <https://doi.org/10.1002/ese3.1203>.
- [9] Mamun A Al, Sohel M, Mohammad N, Haque Sunny MS, Dipta DR, Hossain E. A comprehensive review of the load forecasting techniques using single and hybrid predictive models. *IEEE Access* 2020;8:134911–39. <https://doi.org/10.1109/ACCESS.2020.3010702>.
- [10] Joy TT, Rana S, Gupta S, Venkatesh S. Batch Bayesian optimization using multi-scale search. *Knowl Base Syst* 2020;187:104818. <https://doi.org/10.1016/j.knosys.2019.06.026>.
- [11] Hao Y, Yang W, Yin K. Novel wind speed forecasting model based on a deep learning combined strategy in urban energy systems. *Expert Syst Appl* 2023;219:119636. <https://doi.org/10.1016/j.eswa.2023.119636>.
- [12] Wang J, Yang Z. Ultra-short-term wind speed forecasting using an optimized artificial intelligence algorithm. *Renew Energy* 2021;171:1418–35. <https://doi.org/10.1016/j.renene.2021.03.020>.
- [13] Ahmad T, Zhang D, Huang C. Methodological framework for short-and medium-term energy, solar and wind power forecasting with stochastic-based machine learning approach to monetary and energy policy applications. *Energy* 2021;231:120911. <https://doi.org/10.1016/j.energy.2021.120911>.
- [14] Wang Y, Zhao K, Hao Y, Yao Y. Short-term wind power prediction using a novel model based on butterfly optimization algorithm-variational mode decomposition-long short-term memory. *Appl Energy* 2024;366:123313. <https://doi.org/10.1016/j.apenergy.2024.123313>.
- [15] Karijadi I, Chou S-Y, Dewabharata A. Wind power forecasting based on hybrid CEEMDAN-EWT deep learning method. *Renew Energy* 2023;218:119357. <https://doi.org/10.1016/j.renene.2023.119357>.
- [16] Gong Z, Wan A, Ji Y, Al-Bukhaiti K, Yao Z. Improving short-term offshore wind speed forecast accuracy using a VMD-PE-FCGRU hybrid model. *Energy* 2024;295:131016. <https://doi.org/10.1016/j.energy.2024.131016>.
- [17] Ye F, Brodie J, Miles T, Aziz Ezzat A. AIRU-WRF: a physics-guided spatio-temporal wind forecasting model and its application to the U.S. Mid Atlantic offshore wind energy areas. *Renew Energy* 2024;223:119934. <https://doi.org/10.1016/j.renene.2023.119934>.
- [18] Kazemzadeh M-R, Amjadian A, Amraee T. A hybrid data mining driven algorithm for long term electric peak load and energy demand forecasting. *Energy* 2020;204:117948. <https://doi.org/10.1016/j.energy.2020.117948>.
- [19] Yang K, Zhang X, Luo H, Hou X, Lin Y, Wu J, et al. Predicting energy prices based on a novel hybrid machine learning: comprehensive study of multi-step price forecasting. *Energy* 2024;298:131321. <https://doi.org/10.1016/j.energy.2024.131321>.
- [20] Jogunola O, Adebisi B, Hoang K Van, Tsado Y, Popoola SI, Hammoudeh M, et al. CBLSTM-AE: a hybrid deep learning framework for predicting energy consumption. *Energies* 2022;15:810. <https://doi.org/10.3390/en15030810>.
- [21] Arrieta-Prieto M, Schell KR. Spatio-temporal probabilistic forecasting of wind power for multiple farms: a copula-based hybrid model. *Int J Forecast* 2022;38:300–20. <https://doi.org/10.1016/j.ijforecast.2021.05.013>.
- [22] Aksan F, Suresh V, Janik P, Sikorski T. Load forecasting for the laser metal processing industry using vmd and hybrid deep learning models. *Energies* 2023;16:5381. <https://doi.org/10.3390/en16145381>.
- [23] Xia C, Wang Z. Drivers analysis and empirical mode decomposition based forecasting of energy consumption structure. *J Clean Prod* 2020;254:120107. <https://doi.org/10.1016/j.jclepro.2020.120107>.
- [24] Zhang S, Chen Y, Xiao J, Zhang W, Feng R. Hybrid wind speed forecasting model based on multivariate data secondary decomposition approach and deep learning algorithm with attention mechanism. *Renew Energy* 2021;174:688–704. <https://doi.org/10.1016/j.renene.2021.04.091>.
- [25] Lu Y, Sheng B, Fu G, Luo R, Chen G, Huang Y. Prophet-EEMD-LSTM based method for predicting energy consumption in the paint workshop. *Appl Soft Comput* 2023;143:110447. <https://doi.org/10.1016/j.asoc.2023.110447>.
- [26] Zhang Y, Chen Y. Application of hybrid model based on CEEMDAN, SVD, PSO to wind energy prediction. *Environ Sci Pollut Res* 2022;29:22661–74. <https://doi.org/10.1007/s11356-021-16997-3>.
- [27] Liu S, Xu T, Du X, Zhang Y, Wu J. A hybrid deep learning model based on parallel architecture TCN-LSTM with Savitzky-Golay filter for wind power prediction. *Energy Convers Manag* 2024;302:118122. <https://doi.org/10.1016/j.enconman.2024.118122>.
- [28] Memarzadeh G, Keynia F. Short-term electricity load and price forecasting by a new optimal LSTM-NN based prediction algorithm. *Elec Power Syst Res* 2021;192:106995. <https://doi.org/10.1016/j.epsr.2020.106995>.
- [29] Fekri MN, Patel H, Grolinger K, Sharma V. Deep learning for load forecasting with smart meter data: Online adaptive recurrent neural network. *Appl Energy* 2021;282:116177. <https://doi.org/10.1016/j.apenergy.2020.116177>.
- [30] Zhang Y, Guo H, Sun M, Liu S, Forrest J. A novel grey Lotka–Volterra model driven by the mechanism of competition and cooperation for energy consumption forecasting. *Energy* 2023;264:126154. <https://doi.org/10.1016/j.energy.2022.126154>.
- [31] Jiang Z, Che J, He M, Yuan F. A CGRU multi-step wind speed forecasting model based on multi-label specific XGBoost feature selection and secondary decomposition. *Renew Energy* 2023;203:802–27. <https://doi.org/10.1016/j.renene.2022.12.124>.
- [32] Fu W, Zhang K, Wang K, Wen B, Fang P, Zou F. A hybrid approach for multi-step wind speed forecasting based on two-layer decomposition, improved hybrid DE-HHO optimization and KELM. *Renew Energy* 2021;164:211–29. <https://doi.org/10.1016/j.renene.2020.09.078>.
- [33] Chen J, Fu X, Zhang L, Shen H, Wu J. A novel offshore wind power prediction model based on TCN-DANet-sparse transformer and considering spatio-temporal coupling in multiple wind farms. *Energy* 2024;308:132899. <https://doi.org/10.1016/j.energy.2024.132899>.
- [34] Bentzen LØ, Warakagoda ND, Stenbro R, Engelstad P. Spatio-temporal wind speed forecasting using graph networks and novel transformer architectures. *Appl Energy* 2023;333:120565. <https://doi.org/10.1016/j.apenergy.2022.120565>.
- [35] Zhou H, Li J, Zhang S, Zhang S, Yan M, Xiong H. Expanding the prediction capacity in long sequence time-series forecasting. *Artif Intell* 2023;318:103886. <https://doi.org/10.1016/j.artint.2023.103886>.
- [36] Li S, Jin X, Yuan Y, Zhou X, Chen W, Wang Y-X, et al. Enhancing the locality and breaking the memory bottleneck of transformer on time series forecasting. 2019.
- [37] Lim B, Arik SÖ, Loeff N, Pfister T. Temporal fusion transformers for interpretable multi-horizon time series forecasting. *Int J Forecast* 2021;37:1748–64. <https://doi.org/10.1016/j.ijforecast.2021.03.012>.
- [38] Li D, Tan Y, Zhang Y, Miao S, He S. Probabilistic forecasting method for mid-term hourly load time series based on an improved temporal fusion transformer model. *Int J Electr Power Energy Syst* 2023;146:108743. <https://doi.org/10.1016/j.ijepes.2022.108743>.
- [39] Wu B, Wang L, Zeng Y-R. Interpretable wind speed prediction with multivariate time series and temporal fusion transformers. *Energy* 2022;252:123990. <https://doi.org/10.1016/j.energy.2022.123990>.
- [40] Zhao L, Liu C, Yang C, Liu S, Zhang Y, Li Y. A location-centric transformer framework for multi-location short-term wind speed forecasting. *Energy Convers Manag* 2025;328:119627. <https://doi.org/10.1016/j.enconman.2025.119627>.
- [41] Sun S, Liu Y, Li Q, Wang T, Chu F. Short-term multi-step wind power forecasting based on spatio-temporal correlations and transformer neural networks. *Energy Convers Manag* 2023;283:116916. <https://doi.org/10.1016/j.enconman.2023.116916>.
- [42] Zeng H, Wu B, Fang H, Lin J. Interpretable wind speed forecasting through two-stage decomposition with comprehensive relative importance analysis. *Appl Energy* 2025;392:126015. <https://doi.org/10.1016/j.apenergy.2025.126015>.
- [43] Li C, Shi J. A novel CNN-LSTM-based forecasting model for household electricity load by merging mode decomposition, self-attention and autoencoder. *Energy* 2025;330:136883. <https://doi.org/10.1016/j.energy.2025.136883>.
- [44] Torres ME, Colominas MA, Schlotthauer G, Flandrin P. A complete ensemble empirical mode decomposition with adaptive noise. In: 2011 IEEE int. Conf. Acoust. Speech signal process. IEEE; 2011. p. 4144–7. <https://doi.org/10.1109/ICASSP.2011.5947265>.
- [45] Huang NE, Shen Z, Long SR, Wu MC, Shih HH, Zheng Q, et al. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc R Soc London Ser A Math Phys Eng Sci* 1998;454:903–95. <https://doi.org/10.1098/rspa.1998.0193>.
- [46] Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Comput* 1997;9:1735–80. <https://doi.org/10.1162/neco.1997.9.1735>.
- [47] Ayoub S, Gulzar Y, Reeg FA, Turaev S. Generating image captions using bahdanau attention mechanism and transfer learning. *Symmetry* 2022;14:2681. <https://doi.org/10.3390/sym14122681>.
- [48] Wang K, Qi X, Liu H. A comparison of day-ahead photovoltaic power forecasting models based on deep learning neural network. *Appl Energy* 2019;251:113315. <https://doi.org/10.1016/j.apenergy.2019.113315>.
- [49] Tsegaye S, Sanjeevikumar P, Berling Tjernberg L, Fante KA. Short term load forecasting of electrical power distribution system using enhanced deep neural networks (DNNs). *IEEE Access* 2024. <https://doi.org/10.1109/ACCESS.2024.3432647>. 1–1.
- [50] Zhu L, Laptev N. Deep and confident prediction for time series at uber. In: 2017 IEEE int. Conf. Data Min. Work. IEEE; 2017. p. 103–10. <https://doi.org/10.1109/ICDMW.2017.19>.
- [51] <https://www.dropbox.com/sh/1dczb673bx9xz1/AABII5ePMWdFhAk-dEcRGs1La?dl=0>. Accessed September 2023.
- [52] Australian energy market operator (AEMO). www.aemo.com.au. accessed September 2023.
- [53] <https://frost.met.no/index.html>. Accessed August 2024.

- [54] Azur MJ, Stuart EA, Frangakis C, Leaf PJ. Multiple imputation by chained equations: what is it and how does it work? *Int J Methods Psychiatr Res* 2011;20: 40–9. <https://doi.org/10.1002/mpr.329>.
- [55] Stojković M, Kostić S, Plavšić J, Prohaska S. A joint stochastic-deterministic approach for long-term and short-term modelling of monthly flow rates. *J Hydrol* 2017;544:555–66. <https://doi.org/10.1016/j.jhydrol.2016.11.025>.
- [56] Lago J, Marcjasz G, De Schutter B, Weron R. Forecasting day-ahead electricity prices: a review of state-of-the-art algorithms, best practices and an open-access benchmark. *Appl Energy* 2021;293:116983. <https://doi.org/10.1016/j.apenergy.2021.116983>.
- [57] Giacomini R, White H. Tests of conditional predictive ability. *Econometrica* 2006; 74:1545–78. <https://doi.org/10.1111/j.1468-0262.2006.00718.x>.
- [58] Ben Taieb S, Sorjamaa A, Bontempi G. Multiple-output modeling for multi-step-ahead time series forecasting. *Neurocomputing* 2010;73:1950–7. <https://doi.org/10.1016/j.neucom.2009.11.030>.