

Article

Interval-Valued Multi-Step-Ahead Forecasting of Green Electricity Supply Using Augmented Features and Deep-Learning Algorithms

Tzu-Chi Liu ¹, Chih-Te Yang ² , I-Fei Chen ^{3,*} and Chi-Jie Lu ^{1,4,5,*} ¹ Graduate Institute of Business Administration, Fu Jen Catholic University, New Taipei City 242062, Taiwan² Department of Business Administration, Tamkang University, New Taipei City 251301, Taiwan; 122300@mail.tku.edu.tw³ Department of Management Sciences, Tamkang University, New Taipei City 251301, Taiwan⁴ Department of Information Management, Fu Jen Catholic University, New Taipei City 242062, Taiwan⁵ Artificial Intelligence Development Center, Fu Jen Catholic University, New Taipei City 242062, Taiwan

* Correspondence: 133159@o365.tku.edu.tw (I.-F.C.); 059099@mail.fju.edu.tw (C.-J.L.)

Abstract

Accurately forecasting the interval-valued green electricity (GE) supply is challenging due to the unpredictable and instantaneous nature of its source; yet, reliable multi-step-ahead forecasting is essential for providing the lead time required in operations, resource allocation, and system management. This study proposes an augmented-feature multi-step interval-valued forecasting (AFMIF) scheme that aims to address the challenges in forecasting interval-valued GE supply data by extracting additional features hidden within an interval. Unlike conventional methods that rely solely on original interval bounds, AFMIF integrates augmented features that capture statistical and dynamic properties to reveal hidden patterns. These features include basic interval boundaries and statistical distributions from an interval. Three effective forecasting methods, based on gated recurrent units (GRUs), long short-term memory (LSTM), and a temporal convolutional network (TCN), are constructed under the proposed AFMIF scheme, while the mean ratio of exclusive-or (MRXOR) is used to evaluate the forecasting performance. Two different real datasets of wind-based GE supply data from Belgium and Germany are used as illustrative examples. Empirical results demonstrate that the proposed AFMIF scheme with GRUs can generate promising results, achieving a mean MRXOR of 0.7906 from the Belgium data and 0.9719 from the Germany data for one-step- to three-steps-ahead forecasting. Moreover, the TCN yields an average improvement of 13% across all time steps with the proposed scheme. The results highlight the potential of the AFMIF scheme as an effective alternative approach for accurate multi-step-ahead interval-valued GE supply forecasting that offers practical benefits supporting GE management.

Keywords: green electricity; interval-valued data; multi-step-ahead forecasting; augmented features; deep learning

MSC: 62M10; 68T07; 62P12; 62P30



Academic Editor: Raymond Lee

Received: 28 July 2025

Revised: 31 August 2025

Accepted: 4 September 2025

Published: 6 October 2025

Citation: Liu, T.-C.; Yang, C.-T.; Chen, I.-F.; Lu, C.-J. Interval-Valued Multi-Step-Ahead Forecasting of Green Electricity Supply Using Augmented Features and Deep-Learning Algorithms. *Mathematics* **2025**, *13*, 3202. <https://doi.org/10.3390/math13193202>

Copyright: © 2025 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the Creative Commons

Attribution (CC BY) license

(<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

With the rapid increase in electricity demand and concerns over greenhouse gas emissions from fossil fuels [1–4], renewable energy has become a vital alternative for sustainable

electricity generation [5–7]. However, sources such as wind and solar are inherently variable and unpredictable, creating challenges in ensuring a stable and reliable electricity supply [7,8]. This variability makes the accurate forecasting of green electricity (GE) generation essential for effective operation and planning [9,10]. Given the essential characteristic of the natural resources used for GE generation, the daily actual production output is often unstable, exhibiting a wide range of interval values, which can be represented as interval-valued GE supply data [11]. Interval-valued GE supply data represent GE production records over a day, captured at specific recording intervals (e.g., 15 min per record). This characteristic presents significant challenges for GE supply management. Consequently, the effective forecasting of interval-valued GE supply data is a crucial area of investigation [12–14]. For example, the potential maximum and minimum electricity production can be estimated when forecasting the interval values of GE supply. This information is crucial for power-grid managers to monitor the GE supply and to make informed decisions [14,15]. Consequently, given the inherent complexity of GE data, the development of effective forecasting models and the delivery of useful information to support decision-making are valuable.

Interval-valued forecasting is an important topic in studies concerning GE supply [16,17]. For example, Gupta et al. [18] forecasted interval-valued wind power supply data by combining wavelet and autoregressive integrated moving average (ARIMA) models. Wang et al. [19] employed a multi-objective feature extraction approach to reconstruct decomposed interval-valued wind power data during forecasting. Yang et al. [20] proposed an ensemble-based forecasting system for interval-valued wind power supply data.

However, most existing studies have paid limited attention to multi-step-ahead forecasting for interval-valued GE data [18–20], with those that have often focusing on point forecasts [21,22]. To illustrate, Yu et al. [23] combined convolution layers and long short-term memory (LSTM) for forecasting wind speed 10 min ahead. Nguyen et al. [24] utilized stacked temporal convolutional networks (TCNs) to forecast wind power output values from one to six steps ahead, with each step length being 30 min-based. Tsegaye et al. [25] forecasted power loads with hour- and day-based step lengths using LSTM with population-based adaptive optimization. Given the inherently uncertain nature of GE data, one-step-ahead forecasting provides limited information [16]. For instance, preparing a GE supply often requires lead time for scheduling, coordination, and resource allocation [26]. However, one-step-ahead forecasting offers a narrow window of information for action, providing limited flexibility to plan schedules or manage unexpected fluctuations [12]. In contrast, multi-step-ahead forecasting extends the forecasting horizon, generating more future information that extends the buffer time during operation. This enables managers to gain more informed insights regarding the expected range of the GE supply over longer future horizons, supporting more effective planning such as maintenance or resource allocation [12,16]. Therefore, the accurate multi-step-ahead forecasting of the interval-valued GE supply can benefit GE management.

Furthermore, when forecasting interval-valued GE supply data, the targets are typically represented in either the upper and lower bound or center and radius formats [14]. Either format can express the upper and lower bounds of an interval-valued data point. Most existing studies on forecasting interval-valued GE supply data have primarily considered basic features, such as the time-lagged information of the target variable, as predictor variables directly derived from upper and lower bounds or central and radius data [14]. However, these basic features fail to capture other useful information for the forecasting task, such as the distribution information within an interval-valued data point [27–29]. This information can be described using augmented features such as the quartile values, mean, interquartile range (IQR), skewness, and kurtosis. These augmented features can capture

inherent information from an interval-valued data point and can be incorporated into the forecasting model to enhance the forecasting performance. Few studies have considered augmented features when forecasting interval-valued GE supply data. Jiang et al. [27] demonstrated the potential of utilizing augmented features, but they only considered the quartile values, minimum value, and maximum value for the one-step-ahead forecasting of the regular electricity demand, and not GE supply data. Chang et al. [28] took a similar approach and added the mean, IQR, skewness, and kurtosis as additional augmented features for the one-step-ahead forecasting of the regular electricity supply. However, to the best of our knowledge, no study has simultaneously considered both augmented feature usage and multi-step-ahead forecasting for interval-valued GE supply data. Therefore, addressing this research gap warrants further investigation.

Various methods are utilized for the interval-valued forecasting of GE supply data, including seasonal ARIMA (SARIMA) and deep-learning (DL) algorithms [10,30]. Because interval-valued forecasting involves a pair of target variables (upper and lower bounds or central and radius), it is considered a multi-output task that DL models can handle effectively and conveniently. For example, Niu et al. [31] proposed a framework combining LSTM networks with several decomposition methods to achieve high performance in wind power interval-valued forecasting. Zhou et al. [32] forecasted wind power intervals using LSTM. Wang et al. [33] proposed a framework using a modified scaling approach and efficient feature ranking to improve the performance of gated recurrent unit (GRU) networks in wind power interval-valued forecasting. Wang et al. [34] forecasted wind power intervals with a modified GRU combining two layers of decomposition methods. TCNs have also been used in interval-valued forecasting studies [35,36]. Hu et al. [35] proposed and applied a slightly modified TCN algorithm for wind power interval forecasting, whereas Gan et al. [36] added an adaptive layer to the TCN layer to predict wind speed intervals. Therefore, as interval-valued forecasting is a multi-output task, the commonly used LSTM, GRU, and TCN models are employed in this study.

This study proposes a multi-step-ahead forecasting scheme that incorporate augmented features (base and distribution features) and DL models (GRU, LSTM, and TCN) to forecast the interval-valued GE supply effectively. The proposed scheme is named the Augmented Feature Multi-Step Interval-valued Forecasting (AFMIF) scheme and consists of four steps. First, interval-valued data are generated by preprocessing the original GE supply data, which are records of daily electricity logs within a given time cycle. Second, augmented features are constructed from the interval-valued data that are represented in the upper and lower bound and central and radius formats. Third, LSTM, GRU, and TCN models are constructed, incorporating the augmented and basic features of the interval-valued data, to create three multi-step-ahead interval-valued forecasting models, namely, AFMIF-GRU, AFMIF-LSTM, and AFMIF-TCN, respectively. Finally, these three forecasting models are compared with four other models to evaluate the performance of the proposed AFMIF scheme, and the best model is selected. SARIMA is one of the comparison models used as the benchmark method as it is a commonly used method in GE data forecasting. The mean ratio of exclusive-or (MRXOR) is used to assess the overlap between the actual and estimated intervals. Wind power is a key source of GE and has gained global popularity, with a rapid annual growth rate of 12.6% in wind turbine installations in recent years [9,37–39]. Thus, this study uses wind-power-based GE supply data from Belgium and Germany as illustrative examples to evaluate the performance of the proposed AFMIF scheme. Table 1 provides a comparative overview of the research gap and highlights the contributions of this study relative to the existing literature.

Table 1. Overview of this study compared to existing literatures.

Study	Forecasting Models	Feature Usage	Time-Step	Target
This study	AFMIF incorporates GRU, LSTM, and TCN	Basic and augmented (quartile, minimum, maximum, IQR, mean, kurtosis, and skewness values)	Multi-step	Interval-valued
Jiang et al. [27]	Support vector regression	Basic and augmented (quartile, minimum, and maximum values)	One-step	Interval-valued
Chang et al. [28]	Multivariate adaptive regression splines	Basic and augmented (quartile, minimum, maximum, IQR, mean, kurtosis, and skewness values)		
Gupta et al. [18]	ARIMA incorporates Wavelet	Basic		
Wang et al. [19]	Feature extraction and decomposition method	Basic		
Yang et al. [20]	Ensemble of DL, ARIMA, and decomposition method	Basic		
Wang et al. [33]	Modified scaling and efficient feature ranking improving GRU	Basic		
Wang et al. [34]	GRU with decomposition	Basic		
Hu et al. [35]	Conformalized-quantile-regression-based TCN	Basic		
Tsegaye et al. [25]	Population-based adaptive optimization and LSTM	Basic with external variables	Multi-step	Point forecast
Nahid et al. [21]	LSTM incorporates convolution network with decomposition	Basic		
Pavlov-Kagadejev et al. [22]	Metaheuristics-based LSTM with decomposition	Basic		
Yu et al. [23]	LSTM incorporates convolution network	Basic		
Nguyen et al. [24]	Stacked TCN	Basic		

The remainder of this paper is organized as follows: Section 2 introduces the DL methods and proposed AFMIF scheme. Section 3 presents the processed Belgium and Germany data, descriptive statistics of the augmented features, model results, and robustness evaluation. Finally, Section 4 summarizes the findings.

2. Methodology

This section provides a brief introduction to the LSTM, GRU, and TCN methods, followed by a step-by-step explanation of the proposed AFMIF scheme.

2.1. LSTM

LSTM is an extension of the recurrent neural network that utilizes the concept of retaining information over time to capture long-term dependencies for tasks such as time-series forecasting [40,41]. LSTM retains information via memory cells, each of which is made up of an input gate, a forget gate, and an output gate. The forget gate controls the extent to which previous information should be retained or discarded, and the calculation is shown in Equation (1):

$$f_t = \sigma(w_f \cdot [h_{t-1}, x_t] + b_f), \quad (1)$$

where t is the time period during forecasting; w_f is the weight matrix; h_{t-1} is the output of the previous time period; x_t is the input at the current time period; b_f is the bias; and σ is the sigmoid activation function. The input gate regulates the flow of new information into the cell state that determines updating or retaining new information to overwrite previous information. The calculation is expressed as follows:

$$I_t = \sigma(w_I \cdot [h_{t-1}, x_t] + b_I), \quad (2)$$

where I_t yields a value between 0 and 1, where a value closer to 1 means that the information should be updated; w_I is the weight matrix; and b_I denotes the bias. Finally, the latest output value h_t is generated through the output gate according to Equations (3) and (4):

$$o_t = \sigma(w_o \cdot [h_{t-1}, x_t] + b_o) \quad (3)$$

$$h_t = o_t \times \tanh(C_t), \quad (4)$$

where w_o is the weight matrix; b_o denotes the bias; and C_t is the overwritten information. For further details, please refer to the original study [39].

2.2. GRU

The concept of the GRU is similar to that of LSTM, with several differences in the gate mechanism [42,43]. The GRU is a lightweight modification of LSTM that simplifies the gate mechanism by merging the gates into update and reset gates. The reset gate r_t can be treated as the forget gate of LSTM, with the purpose of controlling the amount of historical information that should be forgotten. The calculation is shown in Equation (5):

$$r_t = \sigma(w_r \cdot [h_{t-1}, x_t]), \quad (5)$$

where w_r is the weight matrix; h_{t-1} is the output of the previous time period; x_t is the input of the current time period; and σ is the sigmoid activation function. The update gate z_t determines the amount of historical information to be carried forward to the current time step. The equation for z_t is expressed as follows:

$$z_t = \sigma(w_z \cdot [h_{t-1}, x_t]), \quad (6)$$

where w_z is the weight matrix. Finally, the output h_t at the current state of time step t can be determined using Equation (7):

$$h_t = (1 - z_t) \times h_{t-1} + z_t \times \tilde{h}_t, \quad (7)$$

where $\tilde{h}_t$ is the potential candidate value with new information to be added at current time step t . The GRU mainly manages the retention and retrieval of input information through reset and update gates for forecasting. For further details, please refer to the original study [41].

2.3. TCN

The TCN is a DL architecture designed for sequential data modeling that is commonly used for time-series forecasting tasks [44,45]. The main mechanism of TCN is the combination of dilated convolution, causal convolution, and residual connection. With these components, the predictions at any time step are conditioned solely on past inputs, preserving the temporal order.

Causal convolution ensures temporal causality by restricting predictions at time step t to depend only on earlier time steps. Its receptive field increases linearly with the number of layers, limiting its ability to capture long-range dependencies in sequential data efficiently. The TCN addresses this limitation by incorporating dilated convolution. Dilated convolution is a variation of standard convolution that introduces gaps between consecutive filter elements, allowing the network to expand its receptive field without increasing the number of parameters, which helps the TCN model to capture long-term dependencies more effectively. Hence, the dilated operation of F on element time step t in a time-series sequence $F(t)$ can be expressed as follows:

$$F(t) = (x \times_d f)(t) = \sum_{i=0}^{k-1} f(i) \cdot x_{t-d \cdot i}, \quad (8)$$

where d is the dilation rate; k is the filter size; and $x_{t-d \cdot i}$ is the past state from the previous convolution. Moreover, to improve model stability and reduce the vanishing gradient effect, residual blocks are incorporated into the TCN. A residual block consists of stacked dilated causal convolution layers, followed by a rectified linear unit (ReLU), normalization layers, and a dropout layer for regularization.

LSTM, the GRU, and the TCN were selected because they are well-established DL models that have demonstrated strong performance in time-series forecasting tasks. LSTM and the GRU belong to the recurrent neural network family and are effective in capturing both the long- and short-term temporal dynamics that are suitable for modeling the evolving behavior of complex interval-valued GE supply data. The TCN, which is a convolution-based model, offers the advantages of parallel processing and stable gradients, making it well-suited for learning the temporal patterns that are hidden within the interval-valued data. Moreover, as mentioned in the Introduction, interval-valued GE supply forecasting is a multi-output task that involves two target variables. The three selected DL models are more effective for handling multi-output tasks, as traditional models often require separate models to be built for each target variable.

2.4. Proposed AFMIF Scheme

Figure 1 presents the proposed AFMIF scheme.

As shown in the figure, the original data consist of GE data that are generated from wind power, comprising electricity logs recorded at a specified minute-based recording cycle, which are collected first. Subsequently, the collected data are preprocessed to generate interval-valued data using the upper and lower bound and central and radius formats. Once the interval-valued data have been prepared, information can be extracted from them to construct augmented features. Two sets of features are constructed in the AFMIF scheme: the basic feature (BF) consisting of time-lagged information of the upper and lower bounds and central and radius; and the distribution feature (DF) consisting of the Q1, median, Q3, IQR, mean, standard deviation (SD), skewness, and kurtosis. The BF and DF represent different information. The BF is the basic information describing the time-lagged information of the upper bound, lower bound, central, and radius data points of the interval-valued data. The DF considers the time-lagged information of the detailed statistics that are directly extracted from the interval-valued data. After obtaining the predictor variables constituting the BF and DF, three different DL algorithms, namely, LSTM, the

GRU, and the TCN, are used to construct models for multi-step-ahead forecasting (i.e., one-step-, two-steps-, and three-steps-ahead) of interval-valued wind power GE supply. As the central and radius format is commonly used in interval-valued forecasting tasks, this study uses this format to form the target variable. Finally, the MRXOR metric is used to evaluate the performance of all constructed models.

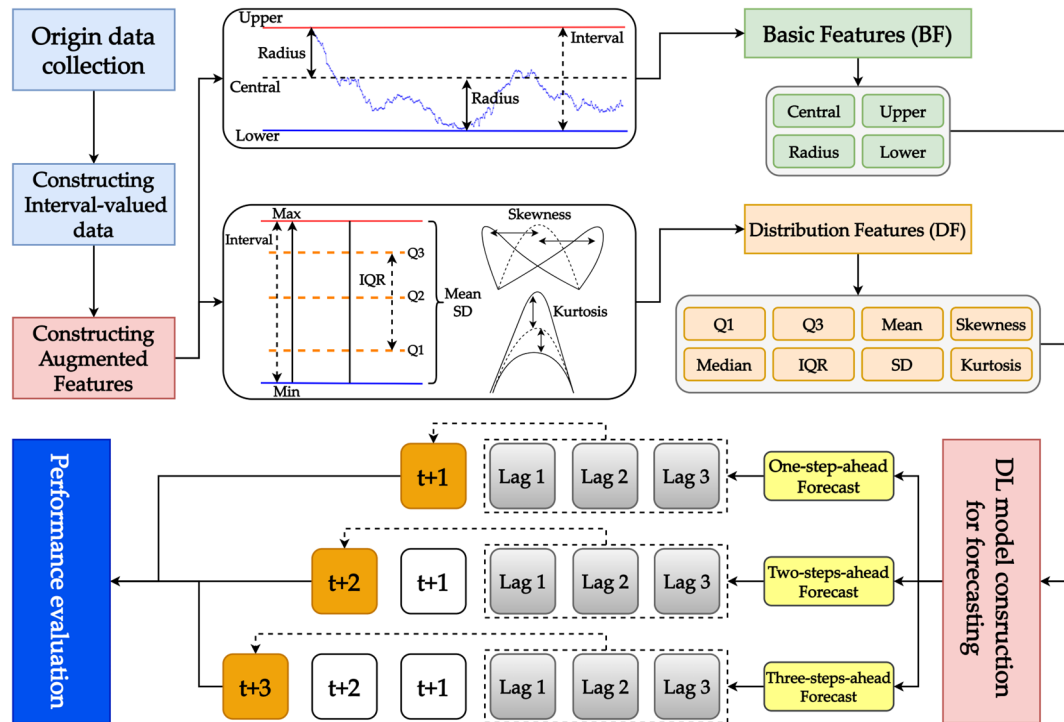


Figure 1. Proposed AFMIF scheme.

The processes of the proposed AFMIF scheme are introduced in the following subsections.

2.4.1. Preprocessing of Interval-Valued Data

The original wind power data systematically represent the electricity produced over time, for which electricity logs are recorded in minute-based cycles per day. Hence, the collected wind power data must first be transformed into interval-valued data. Suppose that a total record (TR) of electricity logs (e) is recorded each day and n days of data are collected. A matrix D_n of size $n \times TR$ can be created and expressed as follows:

$$D_n = \begin{bmatrix} e_{1,1} & \cdots & e_{1,TR} \\ \vdots & e_{i,j} & \vdots \\ e_{n,1} & \cdots & e_{n,TR} \end{bmatrix} = \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_n \end{bmatrix}, \text{ where } i, j \in \mathbb{Z}^+, \quad (9)$$

where $d_i (d_1, d_2, \dots, d_n)$ are vectors of the recorded e logs per day.

Figure 2 shows an example of the first three days of produced electricity data. As shown in the figure, each day contains a total of 96 recorded e logs because the data are recorded under 15 min cycles from 00:00 to 23:45 per day. Thus, D has a matrix size of $n = 3 \times TR = 96$. Then, based on Equation (1), d_1 is a vector consisting of 96 e records ($e_{1,1} = 616.53, e_{1,2} = 608.05, e_{1,3} = 594.07, \dots, e_{1,96} = 210.57$), while d_2 follows the same rule ($e_{2,1} = 213.02, e_{2,2} = 218.46, e_{2,3} = 219.31, \dots, e_{2,96} = 383.3$), and so on. Therefore, the constructed interval-valued data based on the example data are presented in Table 2.

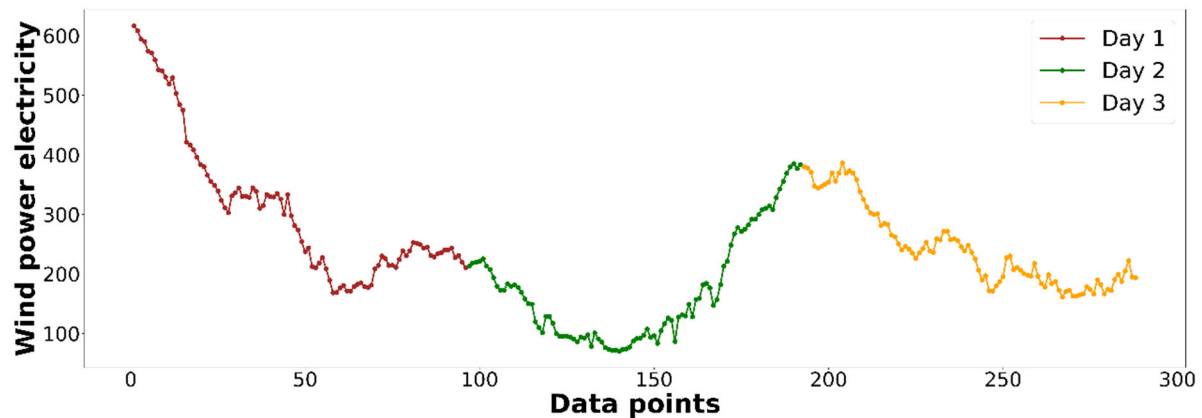


Figure 2. First 3 days of origin wind power data example.

Table 2. First 3 days of interval-valued data example.

Day	Record 1	Record 2	Record 3	...	Record 96
1	616.53	608.05	594.07	...	210.57
2	213.02	218.46	219.31	...	383.3
3	379.72	377.82	370.56	...	193.64

Note: Ellipses indicate intermediate columns omitted for brevity.

2.4.2. Construction of Augmented Features

Once the interval-valued data (matrix D) have been prepared, the augmented features are constructed from the data. The augmented features consist of components of the BF and DF. Both are calculated via the intervals of each day, for which they also form the vectors (d_i) from matrix D .

BF construction:

The BF includes the upper and lower bounds and central and radius. The upper bound expresses the maximum value in an interval, whereas the lower bound expresses the minimum value. They are denoted by Equations (10) and (11), respectively:

$$\text{Upper} = x_i^U = \text{Max}(d_i), \text{ where } i \in \mathbf{Z}^+ = 1 \sim n \quad (10)$$

$$\text{Lower} = x_i^L = \text{Min}(d_i), \text{ where } i \in \mathbf{Z}^+ = 1 \sim n. \quad (11)$$

Following the example data, Upper is the largest value among all 96 records, while Lower is the lowest value among them. Thus, the upper and lower values of the day 1 interval are $\text{Max}(d_1) = 616.53$ and $\text{Min}(d_1) = 168.49$, respectively. The same concept is used to format the upper and lower bounds from the corresponding intervals of each day. After obtaining the upper and lower bounds of each day, they can be further transformed into the central and radius using Equations (12) and (13), respectively:

$$\text{Central} = x_i^C = \frac{1}{2}(x_i^U + x_i^L), \text{ where } i \in \mathbf{Z}^+ = 1 \sim n \quad (12)$$

$$\text{Radius} = x_i^R = \frac{1}{2}(x_i^U - x_i^L), \text{ where } i \in \mathbf{Z}^+ = 1 \sim n. \quad (13)$$

Thus, the central and radius values of day 1 from the example are $x_1^C = \frac{1}{2}(x_1^U + x_1^L) = 392.51$ and $x_1^R = \frac{1}{2}(x_1^U - x_1^L) = 224.02$, respectively.

A pair of upper and lower bound values can be extracted from the interval-valued data of each day; thus, n pairs of upper and lower bound values can be extracted from n days of data. This can be expressed as matrix $\mathbf{X}_i^{UL}$, as shown in Equation (14):

$$\mathbf{X}_i^{UL} = \begin{bmatrix} x_i^U \\ x_i^L \end{bmatrix}^T = \begin{bmatrix} x_1^U & x_1^L \\ \vdots & \vdots \\ x_i^U & x_i^L \\ \vdots & \vdots \\ x_n^U & x_n^L \end{bmatrix} \text{ where } i \in \mathbf{Z}^+. \quad (14)$$

Following the same concept as that of the upper and lower bounds, each pair of central and radius values can be denoted by $\mathbf{X}_i^{CR}$, as shown in Equation (15):

$$\mathbf{X}_i^{CR} = \begin{bmatrix} x_i^C \\ x_i^R \end{bmatrix}^T = \begin{bmatrix} x_1^C & x_1^R \\ \vdots & \vdots \\ x_i^C & x_i^R \\ \vdots & \vdots \\ x_n^C & x_n^R \end{bmatrix}, \text{ where } i \in \mathbf{Z}^+. \quad (15)$$

Both $\mathbf{X}_i^{UL}$ and $\mathbf{X}_i^{CR}$ can provide information to describe the boundary of the interval, which can be viewed as an overall profile expression of the interval.

DF construction:

The structure of an interval can be either symmetric or asymmetric. When the DFs of the mean and median of an interval are equal, with zero skewness, the interval is symmetrically distributed. However, in real-world scenarios, interval-valued data are often asymmetrically distributed and may exhibit complex structures. The DF offers valuable insights into the underlying characteristics of interval-valued data. Hence, utilizing the DF to obtain information that reflects the distribution of an interval can help to enhance the forecasting effectiveness.

The DF consists of the mean, SD, Q1, median, Q3, IQR, skewness, and kurtosis. All of them are constructed using a similar concept to that of the BF in Step 2.1, and calculated via vectors ($\mathbf{d}_i$) from matrix $\mathbf{D}$. Equations (16)–(23) show the DF calculations.

$$\text{Mean} = x_i^M = \text{Average}(\mathbf{d}_i) = \frac{1}{TR} \sum_{j=1}^{TR} e_{i,j}, \quad (16)$$

where $i \in \mathbf{Z}^+ = 1 \sim n$.

$$\text{SD} = x_i^{SD} = \text{Std}(\mathbf{d}_i) = \sqrt{\frac{1}{TR} \sum_{j=1}^{TR} (e_{i,j} - x_i^M)^2}, \quad (17)$$

where $i \in \mathbf{Z}^+ = 1 \sim n$.

$$Q_1 = x_i^{Q_1} = Q_1(\mathbf{d}_i) = \begin{cases} e_{i,j=\lceil \frac{TR}{4} \rceil + 1} & , \text{if } \frac{TR}{4} \notin \mathbf{Z}^+ \\ \frac{1}{2} \left(e_{i,j=\lceil \frac{TR}{4} \rceil} + e_{i,j=\lceil \frac{TR}{4} \rceil + 1} \right) & , \text{if } \frac{TR}{4} \in \mathbf{Z}^+ \end{cases} \quad (18)$$

where $i \in \mathbf{Z}^+ = 1 \sim n$.

$$\text{Median} = x_i^{MD} = \text{Med}(\mathbf{d}_i) = \begin{cases} e_{i,j=(TR+1)/2} & , \text{if } TR \text{ is odd} \\ \frac{1}{2} \left(e_{i,j=\lceil \frac{TR}{2} \rceil} + e_{i,j=\lceil \frac{TR}{2} \rceil + 1} \right) & , \text{if } TR \text{ is even} \end{cases} \quad (19)$$

where $i \in \mathbf{Z}^+ = 1 \sim n$.

$$Q_3 = x_i^{Q_3} = Q_3(d_i) = \begin{cases} e_{i,j=\lceil \frac{3}{4}TR \rceil + 1} & , \text{if } \frac{3}{4}TR \notin \mathbf{Z}^+ \\ \frac{1}{2} \left(e_{i,j=\lceil \frac{3}{4}TR \rceil} + e_{i,j=\lceil \frac{3}{4}TR \rceil + 1} \right) & , \text{if } \frac{3}{4}TR \in \mathbf{Z}^+ \end{cases} \quad (20)$$

where $i \in \mathbf{Z}^+ = 1 \sim n$.

$$IQR = x_i^{IQR} = IQR(d_i) = x_i^{Q_3} - x_i^{Q_1}, \quad (21)$$

where $i \in \mathbf{Z}^+ = 1 \sim n$.

$$\text{Skewness} = x_i^{SK} = 3 \left(x_i^M - x_i^{MD} \right) / x_i^{SD}, \quad (22)$$

where $i \in \mathbf{Z}^+ = 1 \sim n$.

$$\text{Kurtosis} = x_i^{KU} = \text{Kurtosis}(d_i) = \frac{\frac{1}{TR} \sum_{j=1}^{TR} (e_{ij} - x_i^M)^4}{(x_i^{SD})^4} - 3, \quad (23)$$

where $i \in \mathbf{Z}^+ = 1 \sim n$.

Following the example data, the DF values of day 1 are $x_1^M = 310.29$, $x_1^{SD} = 121.99$, $x_1^{Q1} = 222.79$, $x_1^{MD} = 263.97$, $x_1^{Q3} = 345.58$, $x_1^{IQR} = 122.78$, $x_1^{SK} = 1.04$, and $x_1^{KU} = 0.08$.

A set of DFs can be extracted from each day of interval-valued data after construction. Thus, n sets of DFs can be constructed from n days of data, which can be expressed as follows:

$$\mathbf{X}_i^{DF} = \begin{bmatrix} x_i^M \\ x_i^{SD} \\ x_i^{Q1} \\ x_i^{MD} \\ x_i^{Q3} \\ x_i^{IQR} \\ x_i^{SK} \\ x_i^{KU} \end{bmatrix}^T = \begin{bmatrix} x_1^M & x_1^{SD} & x_1^{Q1} & x_1^{MD} & x_1^{Q3} & x_1^{IQR} & x_1^{SK} & x_1^{KU} \\ x_i^M & x_i^{SD} & x_i^{Q1} & x_i^{MD} & x_i^{Q3} & x_i^{IQR} & x_i^{SK} & x_i^{KU} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_n^M & x_n^{SD} & x_n^{Q1} & x_n^{MD} & x_n^{Q3} & x_n^{IQR} & x_n^{SK} & x_n^{KU} \end{bmatrix}, \quad (24)$$

where $i \in \mathbf{Z}^+$.

The augmented features of the BF and DF in each day are constructed from the interval-valued data. Table 3 presents all augmented features extracted from the first three days of interval-valued data as an example.

Table 3. Augmented first 3 days of interval-valued data example.

Day	Central	Radius	Upper	Lower	Mean	SD	Q1	Median	Q3	IQR	Skewness	Kurtosis
1	392.51	224.02	616.53	168.49	310.29	121.99	222.79	263.97	345.58	122.78	1.04	0.08
2	227.54	157.44	384.98	70.10	172.71	89.73	95.47	149.53	219.77	124.31	0.86	−0.35
3	273.54	112.52	386.05	161.02	243.67	67.25	189.29	232.25	281.84	92.55	0.73	−0.69

2.4.3. Training of Multi-Step-Ahead Forecasting Model

After completing the construction of the augmented features, they are utilized for the construction of the forecasting model. The proposed AFMIF scheme uses a direct strategy

for multi-step-ahead forecasting. The direct strategy is a commonly used multi-step-ahead forecasting strategy with a straightforward concept [46,47]. It forecasts the desired target value at h steps ahead from time period t by directly predicting the outcome while ignoring the information within the future time gap [46,47]. This approach builds forecasting models for different time steps ahead separately, allowing each model to focus on learning specific dynamics when making predictions for each time step. Figure 3 presents a visualized example of constructing models under the direct strategy for multi-step-ahead forecasting. The future time gap refers to the time window between time period t and the desired target step ahead during forecasting; for example, $t + 1$ represents forecasting one step ahead. Moreover, as the direct strategy ignores the information within the future time gap, the predictor variables are time-lagged information within the time window of time period t . For example, as shown in the figure, information at time period t and time-lagged information at $t - 1$ and $t - 2$ are used to construct models for multi-step-ahead forecasting.

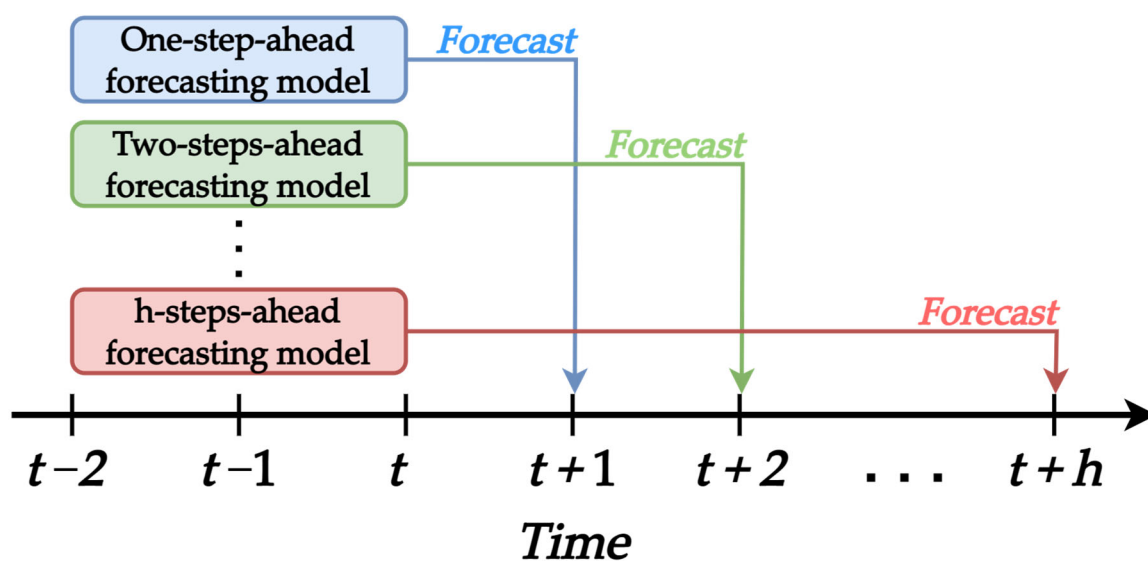


Figure 3. Visualized example of forecasting model built under DS.

The forecasting targets in this study are formed from the central and radius and the forecasting is considered as a multi-output forecasting task. As mentioned in the Introduction, as opposed to traditional methods that must forecast both targets separately, DL models can forecast both targets simultaneously, which is more effective than single-output models. Therefore, let $f()$ be the forecast model and let the time lag be denoted as l . The forecasted central and radius values using the BF and DF as predictors along with the time-lagged information under the direct strategy are expressed as follows:

$$\hat{x}_{t+h}^C = f\left(X_t^{UL}, \dots, X_{t-l+1}^{UL}, X_t^{DF}, \dots, X_{t-l+1}^{DF}\right), \text{ where } t \in i = \mathbf{Z}^+ = \{l, l+1, \dots, n-h\} \quad (25)$$

$$\hat{x}_{t+h}^R = f\left(X_t^{UL}, \dots, X_{t-l+1}^{UL}, X_t^{DF}, \dots, X_{t-l+1}^{DF}\right), \text{ where } t \in i = \mathbf{Z}^+ = \{l, l+1, \dots, n-h\}, \quad (26)$$

where the maximum horizon of future steps considered is three steps ahead ($h = (1, 2, 3)$) and a maximum time-lagged information count of 3 ($l = (1, 2, 3)$) is considered in this study. Hence, three DL models following the approach of the proposed AFMIF scheme are constructed, namely, AFMIF-GRU, AFMIF-LSTM, and AFMIF-TCN. Furthermore, to confirm the effectiveness of the AFMIF scheme, the simple approach of models using only time-lagged information formed from the targets (central and radius) are constructed for comparison; these models are named S-GRU, S-LSTM, and S-TCN. In addition, SARIMA

that is constructed following the simple approach, known as S-SARIMA, is considered for comparison.

All models are built using the expanding window approach based on time when dividing the data into training and testing sets to preserve the chronological order. To ensure fairness and consistency across model comparisons, all three DL models are built using 100 epochs without early stopping, a batch size 64, and the Adam optimizer. ReLU is used as the activation function for all hidden layers, preventing a negative radius, with the linear function for the output layer. Details regarding the DL model structure and corresponding hyperparameters are presented in Supplementary Table S1.

2.4.4. Performance Evaluation

Traditional error metrics evaluate accuracy at the level of single points, measuring the closeness of an estimated point to the corresponding actual point [48]. While these metrics are appropriate for point forecasting, they may be limited or misleading in the interval-valued setting, in which the primary objective is to ensure that the forecasted interval overlaps with the actual interval as closely as possible. Furthermore, the forecasted interval may be wider than the actual interval [48]. Such cases should be considered for appropriate performance evaluation. Thus, the MRXOR metric is used to evaluate the performance of the constructed models when forecasting the GE supply interval. MRXOR considers the overlap of the forecasted and actual intervals [48]. Letting k be the total amount of central and radius data points, the formula for MRXOR is as follows:

$$\text{MRXOR} = \frac{1}{k} \sum_{i=1}^k \frac{\|\mathbf{X}_i^{\text{CR}} \Delta \hat{\mathbf{X}}_i^{\text{CR}}\|}{\|\mathbf{X}_i^{\text{CR}}\|} = \begin{cases} \frac{1}{k} \sum_{i=1}^k \frac{-2(x_i^{\text{R}} - \hat{x}_i^{\text{R}})}{2(x_i^{\text{R}})} & \text{if } \hat{x}_i^{\text{L}} \leq x_i^{\text{L}} \leq x_i^{\text{U}} \leq \hat{x}_i^{\text{U}} \\ \frac{1}{k} \sum_{i=1}^k \frac{2(x_i^{\text{R}} - \hat{x}_i^{\text{R}})}{2(x_i^{\text{R}})} & \text{if } x_i^{\text{L}} \leq \hat{x}_i^{\text{L}} \leq \hat{x}_i^{\text{U}} \leq x_i^{\text{U}} \\ \frac{1}{k} \sum_{i=1}^k \frac{-2(x_i^{\text{C}} - \hat{x}_i^{\text{C}})}{2(x_i^{\text{R}})} & \text{if } x_i^{\text{L}} \leq \hat{x}_i^{\text{L}} \leq x_i^{\text{U}} \leq \hat{x}_i^{\text{U}} \\ \frac{1}{k} \sum_{i=1}^k \frac{2(x_i^{\text{C}} - \hat{x}_i^{\text{C}})}{2(x_i^{\text{R}})} & \text{if } \hat{x}_i^{\text{L}} \leq x_i^{\text{L}} \leq \hat{x}_i^{\text{U}} \leq x_i^{\text{U}} \end{cases}, \quad (27)$$

where $k = n - l - h + 1$; $\hat{x}_i^{\text{L}} = \hat{x}_i^{\text{C}} - \hat{x}_i^{\text{R}}$; $\hat{x}_i^{\text{U}} = \hat{x}_i^{\text{C}} + \hat{x}_i^{\text{R}}$.

The four conditions reflect how the forecasted interval overlaps with an actual interval that is too wide or too narrow, or inclines to the right or inclines to the left. MRXOR is similar to traditional error metrics in that it ranges from zero upwards, with zero representing the ideal case of perfect estimation. Thus, when comparing the MRXOR of different models, better ones will have smaller values closer to zero.

To evaluate the performance of the AFMIF scheme and reliability of the constructed models, robustness evaluation is performed using an expanding window with different training size proportions (60/70/80/90%) to determine whether the model maintains reasonably stable performance across varying training horizons. This approach repeatedly trains the model for the time-series forecasting task, as random data splitting is not suitable; it would break the temporal dependency and disrupt the chronological order. The experiments in this study were implemented in Python version of 3.8.8 [49] and Jupyter Notebook version 6.3.0 [50]. All forecasting models were constructed using the open-source Python packages Scikit-learn (version 0.24.2) [51,52], Keras (version 2.4.3) [53], TensorFlow (version 2.3.0) [54], PyTorch (version 2.3.0) [55], and statsmodels (version 0.14.0) [56].

3. Empirical Results

3.1. Wind Power Data

The wind-power-based GE supply data used to demonstrate the proposed AFMIF scheme in this study were sourced from the Open Power System Data (OPSD) (<https://open-power-system-data.org/>, accessed on 1 September 2024). The OPSD is an open-access platform that provides high-quality, structured, and freely available datasets for power system research and analysis [57]. Two different datasets of GE supply data in Belgium and Germany were used. Germany is one of the largest producers of wind energy in Europe, making it a representative case for large-scale integration. In contrast, Belgium operates at a relatively smaller scale but with consistent recent growth [58,59]. The two countries differ in terms of the integration scale and operating environment, allowing for the applicability of the proposed framework to be evaluated across contexts with different wind power backgrounds.

The dataset for Belgium covers data from January 2013 to May 2017, whereas the Germany dataset covers data from January 2013 to September 2020. Moreover, the data of both were recorded under daily cycles of 15 min per electricity log, resulting in a total of 96 electricity logs recorded in each day. Both datasets follow the introduced process for generating interval-valued data with no missing values, resulting in Belgium and Germany having sample sizes of 1600 and 2830, respectively. For clarity, the datasets from Belgium and Germany are named BE15 and GER15, respectively, in this study. Table 4 presents the descriptive statistics of the augmented features generated from the two datasets. As shown in the table, as the wind power background and scale differ in the two countries, with Germany having a larger scale, the statistical values are significantly different.

Table 4. Descriptive statistics of augmented features from BE15 and GER15 datasets.

Augmented Features		BE15	GER15
Basic Feature (BF)	Upper	427.94 (267.97)	16,666.27 (11,726.67)
	Lower	111.92 (146.02)	6857.93 (7687.09)
	Central	269.93 (193.30)	11,762.10 (9362.54)
	Radius	158.01 (95.93)	4904.17 (3262.81)
Distribution Feature (DF)	Mean	250.23 (208.49)	11,437.16 (9710.70)
	SD	90.86 (61.83)	3009.66 (2187.26)
	Q1	176.98 (188.05)	8961.48 (8878.69)
	Median	242.17 (221.01)	11,234.96 (10,050.65)
	Q3	317.00 (247.23)	13,837.61 (11,018.07)
	IQR	140.03 (115.19)	4876.13 (4104.66)
	Skewness	0.38 (0.71)	0.20 (0.60)
	Kurtosis	−0.40 (1.32)	−0.80 (0.72)

Figures 4 and 5 present the original profile and generated upper and lower bounds of BE15 and GER15. In both figures, the top panel (a) displays the original raw GE supply data profile and the horizontal axis indicates the year; the bottom panel (b) displays the upper and lower bounds of the interval-valued data. The profile of the original data and interval-valued upper and lower bound data appear similar because the interval data were generated by extracting the daily maximum and minimum values from the 96 records, as explained in the Section 2. This process can be viewed as a reduction in data resolution.

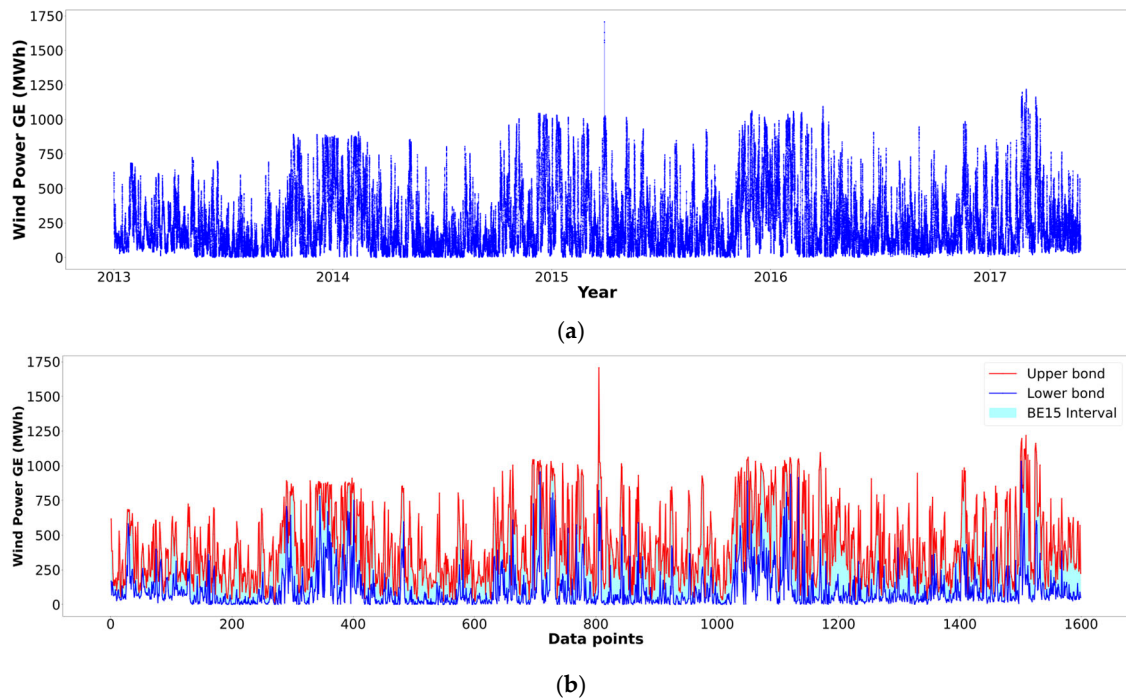


Figure 4. Profile of wind green electricity supply raw and interval-valued data (BE15, Belgium, January 2013 to May 2017) recorded every 15 min per day. (a): Original raw data profile; Panel (b): Upper and lower bounds profile.

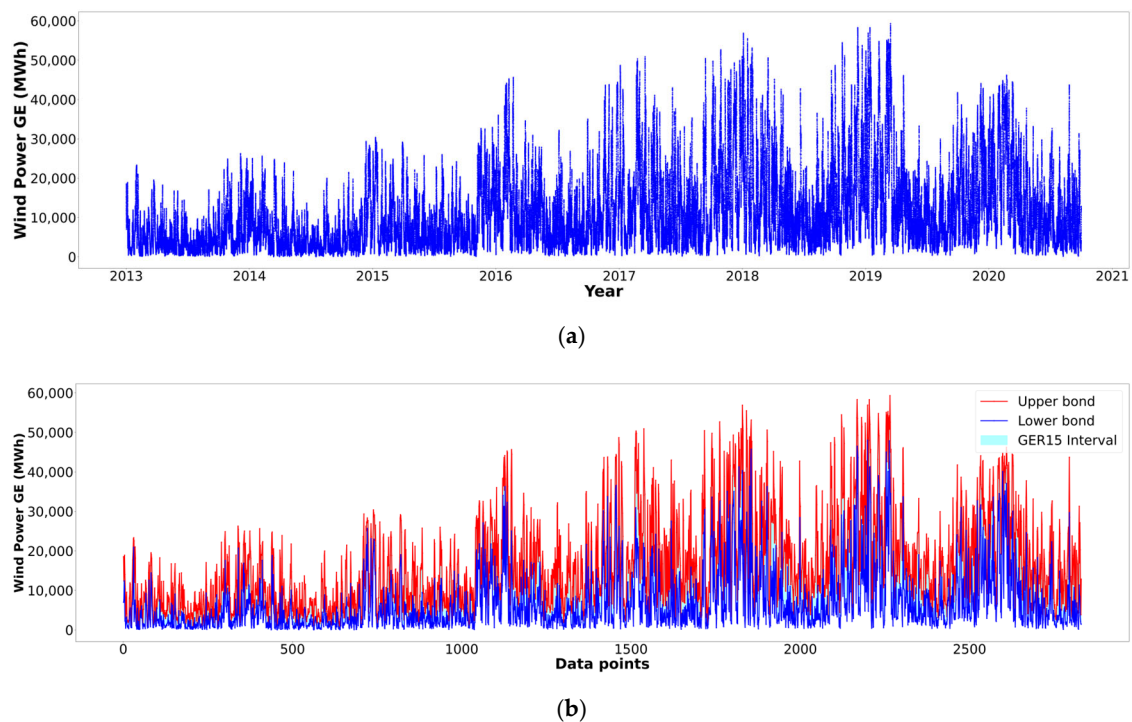


Figure 5. Profile of wind green electricity supply raw and interval-valued data (GER15, Germany, January 2013 to September 2020) recorded every 15 min per day. (a): Original raw data profile; (b): Upper and lower bounds profile.

3.2. Model Results

First, the training size of 80% is discussed. Table 5 presents the forecasting results of models with different approaches for multi-step-ahead forecasting on the BE15 data. As shown in Table 5, in the simple approach, both S-GRU and S-LSTM outperformed

S-SARIMA across all time steps, while TCN-S performed better than SARIMA-S in one-step-ahead forecasting. This confirms the reasonable usage of DL methods and the superiority of the GRU and LSTM compared with the classic SARIMA. Next, in terms of the AFMIF approach when forecasting different time steps ahead, AFMIF-GRU performed the best in one-step- and three-steps-ahead forecasting, while exhibiting a close but inferior performance to AFMIF-LSTM in two-steps-ahead forecasting. This difference is likely owing to the dynamic nature of the data and varying patterns of forecasting uncertainty across different time horizons, resulting in different model reactions. The mean in Table 5 is the average of each model's MRXOR across all time-steps-ahead forecasting. According to the table, AFMIF-GRU was the best model, with a mean MRXOR of 0.7906. In addition, to better demonstrate the overall performance of the proposed AFMIF approach, the overall mean values between S-SARIMA, the simple approach, and the AFMIF approach are compared. As indicated in the table, AFMIF, with an overall mean of 0.8716, performed better than both S-SARIMA and the simple approach. Therefore, AFMIF demonstrated better forecasting effectiveness than the simple approach and the capability for multi-step-ahead forecasting, while AFMI-GRU appeared to be the best-performing model in the BE15 data.

Table 5. MRXOR comparison of models with AFMIF and simple approach using BE15 data.

Approach	Model	One-Step-Ahead	Two-Steps-Ahead	Three-Steps-Ahead	Mean	Overall Mean
Simple	S-SARIMA	1.0215	1.0112	1.0119	1.0149	0.9133
	S-GRU	0.6709	0.8476	0.8801	0.7995	
	S-LSTM	0.7003	0.8436	0.8884	0.8108	
	S-TCN	0.9209	1.0411	1.4269	1.1296	
AFMIF	AFMIF-GRU	0.6604	0.8440	0.8673	0.7906	0.8716
	AFMIF-LSTM	0.6916	0.8432	0.8701	0.8016	
	AFMIF-TCN	0.7025	1.0370	1.3283	1.0226	

To confirm the performance of AFMIF-GRU further, Figure 6 displays the test results for each AFMIF-based model and S-SARIMA. As shown in the figure, the forecasted central and radius were converted into upper and lower bounds and the intervals within the boundaries were filled with color shading for better visualization to compare them with the actual intervals (marked with blue color). In each set of panels, a1 to a3 present the AFMIF-GRU one-step- to three-steps-ahead forecasts and are marked in red; b1 to b3 present the AFMIF-LSTM one-step- to three-steps-ahead forecasts and are marked in green; c1 to c3 present the AFMIF-TCN one-step- to three-steps-ahead forecasts and are marked in orange; and d1 to d3 present the S-SARIMA one-step- to three-steps-ahead forecasts and are marked in brown. The effectiveness of the forecasting results can be assessed by how well the forecasted interval overlaps with and closely follows the profile of the actual interval. For the one-step-ahead forecasting, it can be observed that, with the proposed AFMIF scheme, the forecasted intervals of all three DL methods (panels a1, b1, and c1) aligned well with the actual interval, with AFMIF-GRU demonstrating the best performance in closely following the profile of the actual interval. Moreover, compared with the S-SARIMA one-step-ahead forecasting (panel d1), the sudden spikes of the interval could be captured more effectively with the proposed AFMIF scheme. For the two- and three-steps-ahead forecasting, the drop in performance was noticeable owing to the longer time step ahead, but AFMIF-GRU remained reasonably well-fitted to the actual interval. With these results presented in Figure 6, the performance of AFMIF-GRU, as indicated in Table 5, is further confirmed, and the effectiveness of the proposed AFMIF scheme is demonstrated.

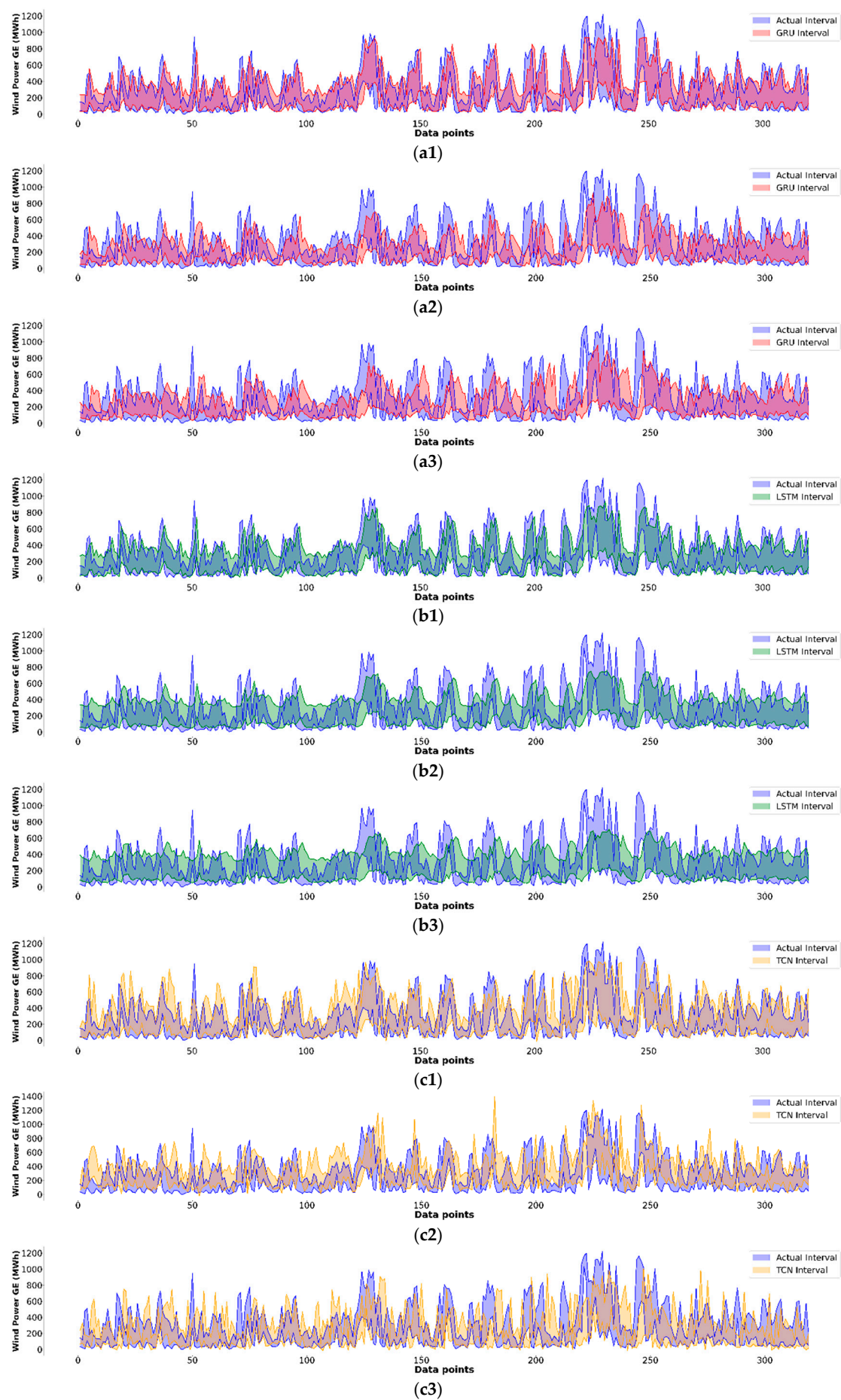


Figure 6. Cont.

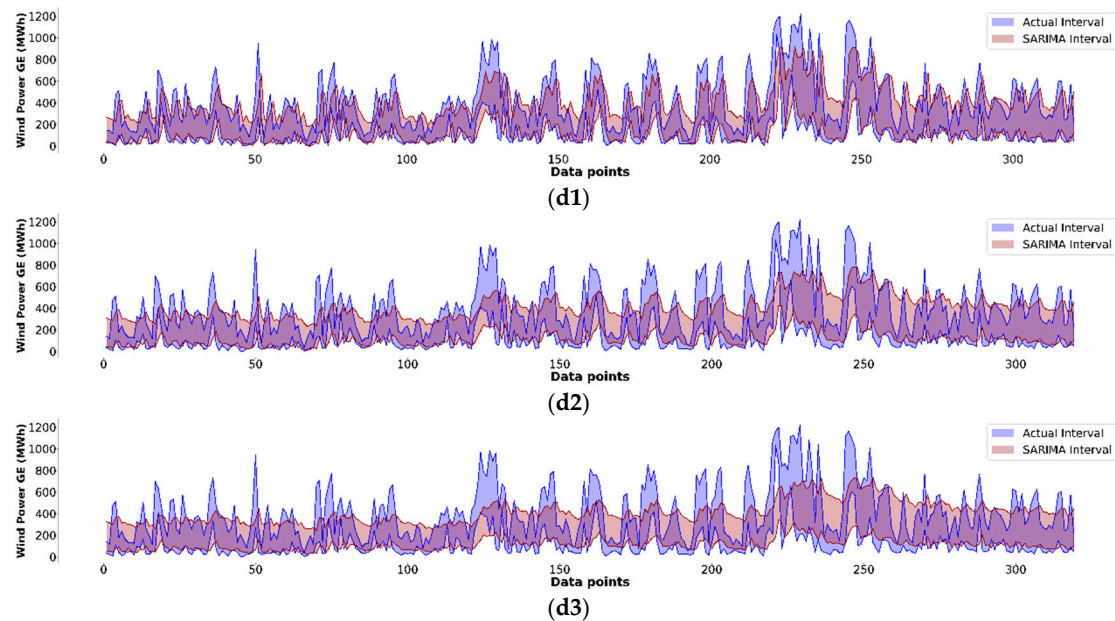


Figure 6. BE15 data interval of testing results from each model. Panels (a1–a3): AFMIF-GRU one-step- to three-steps-ahead forecasting; Panels (b1–b3): AFMIF-LSTM one-step- to three-steps-ahead forecasting; Panels (c1–c3): AFMIF-TCN one-step- to three-steps-ahead forecasting; Panels (d1–d3): S-SARIMA one-step- to three-steps-ahead forecasting.

The effectiveness of AFMIF was also evaluated using the GER15 data with the same settings as those for the BE15 data. Similar circumstances to those found in the BE15 data could be observed in the GER15 data. As shown in Table 6, the DL methods with the simple approach performed better than S-SARIMA, which further validated the DL method usage in this study. AFMIF-GRU also showed superior performance as the top model in the GER15 data, with a mean MRXOR of 0.9719. Moreover, a comparison of the overall means shows that AFMIF maintained better performance than S-SARIMA and the simple approach, with a value of 1.0837. Thus, the promising performance of the proposed AFMIF approach was demonstrated with the BE15 and GER15 data, as well as the capability of the AFMIF scheme for multi-step-ahead forecasting.

Table 6. MRXOR comparison of models with AFMIF and simple approach using GER15 data.

Approach	Model	One-Step-Ahead	Two-Steps-Ahead	Three-Steps-Ahead	Mean	Overall Mean
Simple	S-SARIMA	1.1041	1.3162	1.4678	1.2960	1.1684
	S-GRU	0.7835	1.0557	1.1332	0.9908	
	S-LSTM	0.7861	1.2203	1.1979	1.0681	
	S-TCN	1.0400	1.7147	1.5846	1.4464	
AFMIF	AFMIF-GRU	0.7554	1.0402	1.1201	0.9719	1.0837
	AFMIF-LSTM	0.7737	1.1966	1.1785	1.0496	
	AFMIF-TCN	0.9249	1.3383	1.4258	1.2297	

Comparing the DL methods with the AFMIF approach in both datasets at different time-steps-ahead forecasting according to the information in Table 7, the superiority of AFMIF-GRU is clearly demonstrated. By averaging the MRXOR values in each dataset, AFMIF-GRU achieved the best MRXOR value across all steps ahead, with the one-step-ahead forecasting being the lowest (mean MRXOR = 0.7079). The decreasing performance

in the larger time steps ahead was noticeable in both datasets, but remained reasonable for all models. However, the performance of models using the GER15 data was slightly inferior to those using BE15, which may be owing to the natural structure of the actual data. Nonetheless, the performance was consistent among all methods, which shows the reliability of the proposed AFMIF scheme. Overall, AFMIF-GRU exhibited the best performance in both datasets for multi-step-ahead interval-valued forecasting and the promising performance of the AFMIF scheme was confirmed.

Table 7. Model MRXORs with AFMIF approach comparison between BE15 and GER15 data for multi-step-ahead forecasting.

<i>h</i> -Steps-Ahead	Data	AFMIF-GRU	AFMIF-LSTM	AFMIF-TCN
One-step-ahead	BE15	0.6604	0.6916	0.7025
	GER15	0.7554	0.7737	0.9249
	Mean	0.7079	0.7327	0.8137
Two-steps-ahead	BE15	0.8440	0.8432	1.0370
	GER15	1.0402	1.1966	1.3383
	Mean	0.9421	1.0199	1.1877
Three-steps-ahead	BE15	0.8673	0.8701	1.3283
	GER15	1.1201	1.1785	1.4258
	Mean	0.9937	1.0243	1.3771

The promising potential of the AFMIF scheme could further be demonstrated via the improvement rate that was obtained from the performance of both the BE15 and GER15 datasets for each DL model. The improvement rate was calculated by determining the mean difference between the AFMIF and simple approaches, followed by calculating the ratio of improvement achieved by the AFMIF approach over the simple approach. To compare the overall improvement, the results from both datasets were aggregated by calculating their mean. Equation (28) shows the calculation of the improvement rate:

$$\text{Improvement} = \frac{\text{Simple} - \text{AFMIF}}{\text{Simple}} \times 100\%, \quad (28)$$

where a positive value indicates a reduction in the error. The proposed AFMIF scheme performed better than the simple approach. To illustrate, the mean MRXORs of S-GRU and AFMIF-GRU for the one-step-ahead forecasting using the two datasets were 0.7272 ($\frac{1}{2}(0.6709 + 0.7838)$) and 0.7079 ($\frac{1}{2}(0.6604 + 0.7554)$), respectively. The yielded improvement rate was 2.65% ($100 \times (0.7272 - 0.7079)/0.7079$), where a 2.65% reduction in error was observed when incorporating the AFMIF approach. The improvement rate of each DL model is presented in Figure 7, where bar colors of blue, orange, and red correspond to one-step-, two-steps-, and three-steps-ahead forecasting, respectively. According to Figure 7, the performance of each DL method was enhanced for the one-step-, two-steps-, and three-steps-ahead forecasting, especially for the TCN. Although the TCN was less effective than the GRU or LSTM (as shown in Table 7), a significant improvement in the TCN could be observed when incorporating the AFMIF scheme.

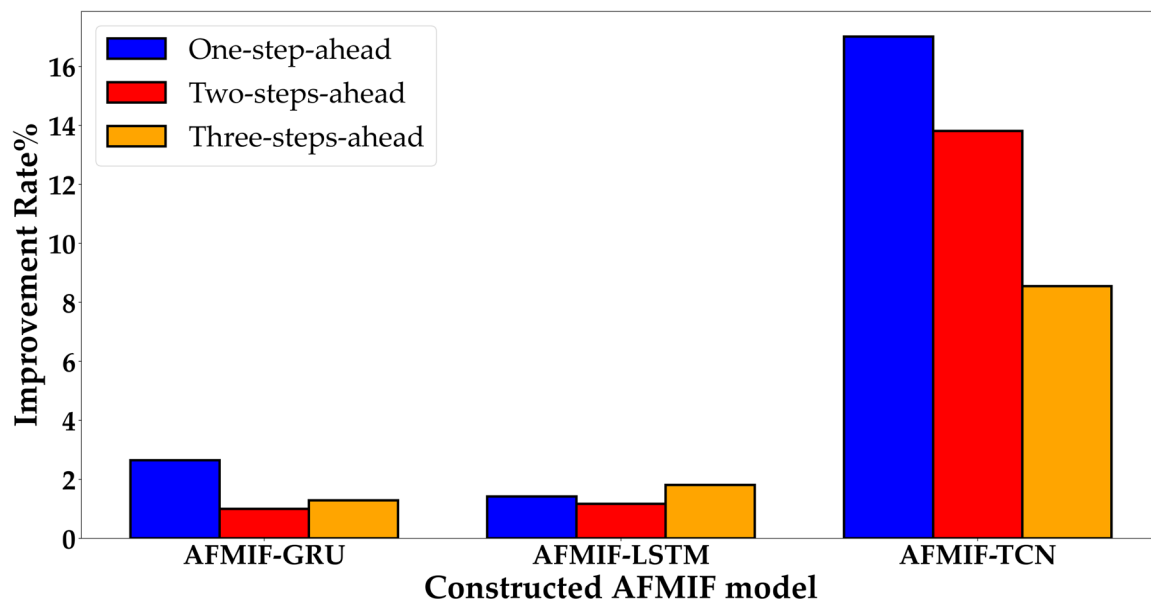


Figure 7. Improvement rate of DL models with AFMIF scheme. (Both BE15 and GER15 results from corresponding simple or AFMIF approach are aggregated when calculating the overall improvement rate).

3.3. Robustness Evaluation

According to the experiments, AFMIF-GRU was the best model in this study. To ensure that the performance of the GRU was stable and the findings of the experiment were reliable, a robustness evaluation was performed on AFMIF-GRU, S-GRU, and S-SARIMA with both the BE15 and GER15 data. For the robustness evaluation, the model was repeatedly constructed using the same process (as mentioned in the third paragraph) with different proportions of training data. The robustness evaluation for the GRU using the BE15 data is presented in Table 8. The MRXORs of each model for the one-step-, two-steps-, and three-steps-ahead forecasting under various training data proportions are shown. The MRXOR values remained consistent among the models, with no dramatic changes in performance. Comparing the mean in each step ahead, AFMIF-GRU with mean MRXORs of 0.6925, 0.8889, and 0.9286 for one-step-, two-steps-, and three-steps-ahead forecasting consistently outperformed both S-SARIMA and S-GRU. Moreover, the overall mean, calculated as the average of all MRXOR values from each model, demonstrated the superiority of AFMIF-GRU (0.8367).

Using the same concept as that for BE15, the robustness evaluation using the GER15 data is presented in Table 9. The GRU with both approaches consistently performed better than S-SARIMA in multi-step-ahead forecasting. The performance of AFMIF-GRU varied slightly with different proportions of training data, but it was acceptable as similar circumstances appeared in S-GRU. With mean MRXORs of 0.8292, 1.1632, and 1.2568 in the one-step-, two-steps-, and three-steps-ahead forecasting, AFMIF-GRU consistently showed superior performance over S-GRU. The overall mean of 1.0831 of AFMIF-GRU also confirmed the effectiveness of the AFMIF approach. Hence, according to the robustness evaluation, the findings of this study are consistent and reliable.

The superior performance of the AFMIF-GRU model in accurately forecasting the upper and lower bounds of the GE supply across multiple time steps can provide additional useful information to support GE management. In real-world GE supply operations, such interval forecasts enable grid operators and energy planners to assess not only the expected production but also the potential variability within a future time window. This allows for more informed decision-making regarding load balancing, reserve allocation, and energy dispatch planning. For example, knowing the likely range of the GE output different

steps ahead provides critical lead times for preparing backup generation, adjusting storage systems, or scheduling maintenance, thereby improving reliability and reducing the risk of system failure or uneven supply. In addition, accurate multi-step-ahead forecasting can improve green energy management and GE supply reliability, potentially reducing the reliance on fossil-based power plants.

Table 8. Robustness evaluation of S-SARIMA, S-GRU, and AFMIF-GRU using BE15 data.

Model	Train%	One-Step-Ahead	Two-Steps-Ahead	Three-Steps-Ahead	Overall Mean
S-SARIMA	60%	1.1254	1.3212	1.3950	1.1691
	70%	1.1719	1.2513	1.2658	
	80%	1.0215	1.0112	1.0119	
	90%	1.1020	1.1708	1.1807	
	Mean	1.1052	1.1886	1.2134	
S-GRU	60%	0.7062	0.9250	0.9423	0.8445
	70%	0.7115	0.9058	0.9482	
	80%	0.6709	0.8476	0.8801	
	90%	0.7253	0.9100	0.9606	
	Mean	0.7035	0.8971	0.9328	
AFMIF-GRU	60%	0.6944	0.9230	0.9411	0.8367
	70%	0.6939	0.8872	0.9464	
	80%	0.6604	0.8440	0.8673	
	90%	0.7212	0.9014	0.9596	
	Mean	0.6925	0.8889	0.9286	

Table 9. Robustness evaluation of S-SARIMA, S-GRU, and AFMIF-GRU using GER15 data.

Model	Train%	One-Step-Ahead	Two-Steps-Ahead	Three-Steps-Ahead	Overall Mean
S-SARIMA	60%	1.1716	1.3963	1.5993	1.3303
	70%	1.1202	1.3030	1.4434	
	80%	1.1041	1.3162	1.4678	
	90%	1.2028	1.3542	1.4847	
	Mean	1.1497	1.3424	1.4988	
S-GRU	60%	0.9210	1.3132	1.4423	1.1004
	70%	0.8710	1.2485	1.3128	
	80%	0.7835	1.0557	1.1332	
	90%	0.8147	1.1237	1.1854	
	Mean	0.8476	1.1853	1.2684	
AFMIF-GRU	60%	0.9169	1.2747	1.4267	1.0831
	70%	0.8433	1.2257	1.3031	
	80%	0.7554	1.0402	1.1201	
	90%	0.8013	1.1122	1.1772	
	Mean	0.8292	1.1632	1.2568	

Although the proposed AFMIF scheme has demonstrated promising results, some limitations remain. While the results showed that incorporating augmented features could benefit the model performance, further efforts are required to improve the model forecasting capability on larger horizons. Validation on both the Belgium and Germany datasets supported the consistency of the proposed scheme, but its versatility may be partially constrained when it is applied to other renewable energy sources such as solar or hydro power, as their data structures differ from that of wind power and may require adaptations to the scheme. Overall, the inherent complexity of interval-valued GE data and the challenge of multi-step-ahead forecasting over extended horizons remain significant. Building on the empirical findings of this study, future research will focus on refining the augmented features and improving the model performance for larger forecasting horizons.

4. Conclusions

This study has proposed the AFMIF scheme for the multi-step-ahead forecasting of interval-valued GE supply data by constructing augmented features such as the BF and DF that represent information on the statistical and dynamic properties hidden within the interval-valued data. Using the actual generated wind power GE data of BE15 and GER15, the proposed AFMIF-GRU showed promising results in multi-step-ahead forecasting, outperforming the traditional SARIMA and simple approaches. The promising model provides accurate estimations of GE supply intervals across different future time horizons, which can enhance decision-making and resource allocation in GE management. This also improves the effectiveness in handling unexpected events, as the forecasted GE intervals serve as early warnings, providing managers with better-informed expectations and more time to react and prepare accordingly. In addition, improved GE management can increase the reliability of green energy, encouraging its adoption. This, in turn, may reduce the reliance on fossil-based power plants and contribute to sustainability. In summary, augmented feature usage can enhance the performance of multi-step-ahead interval-valued forecasting and shows promising potential, as demonstrated with the proposed AFMIF scheme.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/math13193202/s1>, Table S1: Network structure and hyperparameters for each DL model.

Author Contributions: Conceptualization, T.-C.L. and C.-J.L.; data curation, T.-C.L.; formal analysis, T.-C.L., C.-T.Y., I.-F.C. and C.-J.L.; methodology, T.-C.L., C.-T.Y. and C.-J.L.; project administration, C.-J.L.; software, T.-C.L.; supervision, C.-T.Y. and I.-F.C.; writing—original draft, T.-C.L.; writing—review and editing, C.-T.Y., I.-F.C. and C.-J.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was partially supported by the National Science and Technology Council, Taiwan (NSTC112-2221-E-030-011-MY3).

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors upon reasonable request.

Acknowledgments: The authors would like to express their sincere appreciation to Shou-Ming Chen, for his valuable assistance in the data collection, preprocessing, and manuscript formatting and editing stages of this study. His meticulous attention to detail, initiative in identifying relevant data sources, ability to organize and structure complex datasets, and support in ensuring the clarity and consistency of the manuscript significantly contributed to the quality of the research.

Conflicts of Interest: The authors declare that this study was conducted in the absence of any commercial or financial relationships that could be construed as potential conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

GE	Green electricity
ARIMA	Autoregressive integrated moving average
IQR	Interquartile range
SARIMA	Seasonal autoregressive integrated moving average
DL	Deep learning
LSTM	Long short-term memory
GRU	Gated recurrent unit
TCN	Temporal convolutional network
AFMIF	Augmented Feature Multi-Step Interval-valued Forecasting
ReLU	Rectified linear unit
BF	Basic feature
DF	Distribution feature
SD	Standard deviation
MRXOR	Mean ratio of exclusive-or
OPSD	Open power system data
BE15	Wind power dataset recorded under daily 15 min cycle from Belgium
GER15	Wind power dataset recorded under daily 15 min cycle from Germany

References

1. Riaz, A.A.; Hussain, G.; Iqbal, A.; Esat, V.; Alkahtani, M.; Khan, A.M.; Ullah, N.; Xiao, M.; Khan, S. Energy consumption, carbon emissions, product cost, and process time in incremental sheet forming process: A holistic review from sustainability perspective. *Proc. Inst. Mech. Eng. Part B J. Eng. Manuf.* **2022**, *236*, 1683–1705. [\[CrossRef\]](#)
2. Dyrstad, J.M.; Skonhoft, A.; Christensen, M.Q.; Ødegaard, E.T. Does economic growth eat up environmental improvements? Electricity production and fossil fuel emission in OECD countries 1980–2014. *Energy Policy* **2019**, *125*, 103–109. [\[CrossRef\]](#)
3. Hu, X.; Tan, W.; Xie, Y.; Yang, W.; Zeng, Z.; Huang, Y.; Xiao, D.; Chi, Y.; Cheng, R. Operation and Evaluation of Digitalized Retail Electricity Markets under Low-Carbon Transition: Recent Advances and Challenges. *Front. Energy Res.* **2023**, *11*, 1324450. [\[CrossRef\]](#)
4. Ghiassi-Farrokhfal, Y.; Ketter, W.; Collins, J. Making green power purchase agreements more predictable and reliable for companies. *Decis. Support Syst.* **2021**, *144*, 113514. [\[CrossRef\]](#)
5. Thompson, S. Strategic Analysis of the Renewable Electricity Transition: Power to the World without Carbon Emissions? *Energies* **2023**, *16*, 6183. [\[CrossRef\]](#)
6. Davis, M.; Moronkeji, A.; Ahiduzzaman, M.; Kumar, A. Assessment of renewable energy transition pathways for a fossil fuel-dependent electricity-producing jurisdiction. *Energy Sustain. Dev.* **2020**, *59*, 243–261. [\[CrossRef\]](#)
7. Farghali, M.; Osman, A.I.; Chen, Z.; Abdelhaleem, A.; Ihara, I.; Mohamed, I.M.; Yap, P.S.; Rooney, D.W. Social, environmental, and economic consequences of integrating renewable energies in the electricity sector: A review. *Environ. Chem. Lett.* **2023**, *21*, 1381–1418. [\[CrossRef\]](#)
8. Mitra, M.; Singha, N.R.; Chattopadhyay, P.K. Review on renewable energy potential and capacities of South Asian countries influencing sustainable environment: A comparative assessment. *Sustain. Energy Technol. Assess.* **2023**, *57*, 103295. [\[CrossRef\]](#)
9. Deng, X.; Shao, H.; Hu, C.; Jiang, D.; Jiang, Y. Wind power forecasting methods based on deep learning: A survey. *Comput. Model. Eng. Sci.* **2020**, *122*, 273. [\[CrossRef\]](#)
10. Sri Revathi, B. A survey on advanced machine learning and deep learning techniques assisting in renewable energy generation. *Environ. Sci. Pollut. Res.* **2023**, *30*, 93407–93421. [\[CrossRef\]](#)
11. Dou, Y.; Tan, S.; Xie, D. Comparison of machine learning and statistical methods in the field of renewable energy power generation forecasting: A mini review. *Front. Energy Res.* **2023**, *11*, 1218603. [\[CrossRef\]](#)
12. Kath, C.; Ziel, F. Conformal prediction interval estimation and applications to day-ahead and intraday power markets. *Int. J. Forecast.* **2021**, *37*, 777–799. [\[CrossRef\]](#)
13. 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. [\[CrossRef\]](#)
14. Wang, P.; Gurmani, S.H.; Tao, Z.; Liu, J.; Chen, H. Interval time series forecasting: A systematic literature review. *J. Forecast.* **2024**, *43*, 249–285. [\[CrossRef\]](#)

15. Lee, H.; Kim, N.W.; Lee, J.G.; Lee, B.T. Uncertainty-aware forecast interval for hourly PV power output. *IET Renew. Power Gener.* **2019**, *13*, 2656–2664. [[CrossRef](#)]
16. Wang, Y.; Zou, R.; Liu, F.; Zhang, L.; Liu, Q. A review of wind speed and wind power forecasting with deep neural networks. *Appl. Energy* **2021**, *304*, 117766. [[CrossRef](#)]
17. Tsai, W.C.; Hong, C.M.; Tu, C.S.; Lin, W.M.; Chen, C.H. A review of modern wind power generation forecasting technologies. *Sustainability* **2023**, *15*, 10757. [[CrossRef](#)]
18. Gupta, A.; Kumar, A.; Boopathi, K. Intraday wind power forecasting employing feedback mechanism. *Electr. Power Syst. Res.* **2021**, *201*, 107518. [[CrossRef](#)]
19. Wang, J.; Lu, Y.; Li, Q. An interpretable interval-valued wind power prediction system based on multi-objective feature extraction and base model selection with dynamic ensemble. *Swarm Evol. Comput.* **2025**, *96*, 101977. [[CrossRef](#)]
20. Yang, W.; Zang, X.; Wu, C.; Hao, Y. A new multi-objective ensemble wind speed forecasting system: Mixed-frequency interval-valued modeling paradigm. *Energy* **2024**, *304*, 131963. [[CrossRef](#)]
21. Nahid, F.A.; Ongsakul, W.; Manjiparambil, N.M.; Singh, J.G.; Roy, J. Mode decomposition-based short-term multi-step hybrid solar forecasting model for microgrid applications. *Electr. Eng.* **2024**, *106*, 3349–3380. [[CrossRef](#)]
22. Pavlov-Kagadejev, M.; Jovanovic, L.; Bacanin, N.; Deveci, M.; Zivkovic, M.; Tuba, M.; Strumberger, I.; Pedrycz, W. Optimizing long-short-term memory models via metaheuristics for decomposition aided wind energy generation forecasting. *Artif. Intell. Rev.* **2024**, *57*, 45.
23. Yu, M.; Tao, B.; Li, X.; Liu, Z.; Xiong, W. Local and Long-range Convolutional LSTM Network: A novel multi-step wind speed prediction approach for modeling local and long-range spatial correlations based on ConvLSTM. *Eng. Appl. Artif. Intell.* **2024**, *130*, 107613. [[CrossRef](#)]
24. Nguyen, H.K.M.; Phan, Q.D.; Wu, Y.K.; Phan, Q.T. Multi-step wind power forecasting with stacked temporal convolutional network (S-TCN). *Energies* **2023**, *16*, 3792. [[CrossRef](#)]
25. Tsegaye, S.; Padmanaban, S.; Tjernberg, L.B.; Fante, K.A. Short-Term Load Forecasting for Electrical Power Distribution Systems Using Enhanced Deep Neural Networks. *IEEE Access* **2024**, *12*, 186856–186871. [[CrossRef](#)]
26. Taloba, A.I.; Rayan, A. Machine learning based on reliable and sustainable electricity supply from renewable energy sources in the agriculture sector. *J. Radiat. Res. Appl. Sci.* **2025**, *18*, 101282. [[CrossRef](#)]
27. Jiang, P.; Li, R.; Liu, N.; Gao, Y. A novel composite electricity demand forecasting framework by data processing and optimized support vector machine. *Appl. Energy* **2020**, *260*, 114243. [[CrossRef](#)]
28. Chang, T.J.; Lee, S.; Lee, J.; Lu, C.J. An interval-valued time series forecasting scheme with probability distribution features for electric power generation prediction. *IEEE Access* **2022**, *10*, 6417–6429. [[CrossRef](#)]
29. Tian, J.; Ooka, R.; Lee, D. Multi-scale solar radiation and photovoltaic power forecasting with machine learning algorithms in urban environment: A state-of-the-art review. *J. Clean. Prod.* **2023**, *426*, 139040. [[CrossRef](#)]
30. Ying, C.; Wang, W.; Yu, J.; Li, Q.; Yu, D.; Liu, J. Deep learning for renewable energy forecasting: A taxonomy, and systematic literature review. *J. Clean. Prod.* **2023**, *384*, 135414.
31. Niu, D.; Sun, L.; Yu, M.; Wang, K. Point and interval forecasting of ultra-short-term wind power based on a data-driven method and hybrid deep learning model. *Energy* **2022**, *254*, 124384. [[CrossRef](#)]
32. Zhou, M.; Wang, B.; Guo, S.; Watada, J. Multi-objective prediction intervals for wind power forecast based on deep neural networks. *Inf. Sci.* **2021**, *550*, 207–220. [[CrossRef](#)]
33. Wang, J.; Li, Q.; Zhang, H.; Wang, Y. A deep-learning wind speed interval forecasting architecture based on modified scaling approach with feature ranking and two-output gated recurrent unit. *Expert. Syst. Appl.* **2023**, *211*, 118419. [[CrossRef](#)]
34. Wang, M.; Ying, F.; Nan, Q. Refined offshore wind speed prediction: Leveraging a two-layer decomposition technique, gated recurrent unit, and kernel density estimation for precise point and interval forecasts. *Eng. Appl. Artif. Intell.* **2024**, *133*, 108435. [[CrossRef](#)]
35. Hu, J.; Luo, Q.; Tang, J.; Heng, J.; Deng, Y. Conformed temporal convolutional quantile regression networks for wind power interval forecasting. *Energy* **2022**, *248*, 123497. [[CrossRef](#)]
36. Gan, Z.; Li, C.; Zhou, J.; Tang, G. Temporal convolutional networks interval prediction model for wind speed forecasting. *Electr. Power Syst. Res.* **2021**, *191*, 106865. [[CrossRef](#)]
37. Depci, T.; İnci, M.; Savrun, M.M.; Büyük, M. A review on wind power forecasting regarding impacts on the system operation, Technical Challenges, and Applications. *Energy Technol.* **2022**, *10*, 2101061. [[CrossRef](#)]
38. Zhao, L.; Nazir, M.S.; Nazir, H.M.J.; Abdalla, A.N. A review on proliferation of artificial intelligence in wind energy forecasting and instrumentation management. *Environ. Sci. Pollut. Res.* **2022**, *29*, 43690–43709. [[CrossRef](#)]
39. Bouddou, R.; Benhamida, F.; Haba, M.; Belgacem, M.; Meziane, M.A. Simulated Annealing Algorithm for Dynamic Economic Dispatch Problem in the Electricity Market Incorporating Wind Energy. *Ingén. Syst. Inf.* **2020**, *25*, 719–727. [[CrossRef](#)]
40. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [[CrossRef](#)]

41. Sherstinsky, A. Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. *Phys. D Nonlinear Phenom.* **2020**, *404*, 132306. [CrossRef]
42. Cho, K.; van Merriënboer, B.; Gulcehre, C.; Bahdanau, D.; Bougares, F.; Schwenk, H.; Bengio, Y. Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, Doha, Qatar, 25–29 October 2014; pp. 1724–1734. [CrossRef]
43. Fantini, D.G.; Silva, R.N.; Siqueira, M.B.B.; Pinto, M.S.S.; Guimarães, M.; Junior, A.B. Wind speed short-term prediction using recurrent neural network GRU model and stationary wavelet transform GRU hybrid model. *Energy Convers. Manag.* **2024**, *308*, 118333. [CrossRef]
44. Lea, C.; Flynn, M.D.; Vidal, R.; Reiter, A.; Hager, G.D. Temporal Convolutional Networks for Action Segmentation and Detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 156–165.
45. Lara-Benítez, P.; Carranza-García, M.; Luna-Romera, J.M.; Riquelme, J.C. Temporal convolutional networks applied to energy-related time series forecasting. *Appl. Sci.* **2020**, *10*, 2322. [CrossRef]
46. Taieb, S.B.; Bontempi, G.; Atiya, A.F.; Sorjamaa, A.A. Review and comparison of strategies for multi-step ahead time series forecasting based on the NN5 forecasting competition. *Expert Syst. Appl.* **2012**, *39*, 7067–7083. [CrossRef]
47. Suradhaniwar, S.; Kar, S.; Durbha, S.S.; Jagarlapudi, A. Time Series Forecasting of Univariate Agrometeorological Data: A Comparative Performance Evaluation via One-Step and Multi-Step Ahead Forecasting Strategies. *Sensors* **2021**, *21*, 2430. [CrossRef] [PubMed]
48. Hsu, H.L.; Wu, B. Evaluating forecasting performance for interval data. *Comput. Math. Appl.* **2008**, *56*, 2155–2163. [CrossRef]
49. Van Rossum, G.; Drake, F.L. *Python 3 Reference Manual*; CreateSpace: Scotts Valley, CA, USA, 2009.
50. Kluyver, T.; Ragan-Kelley, B.; Pérez, F.; Granger, B.; Bussonnier, M.; Frederic, J.; Kelley, K.; Hamrick, J.; Grout, J.; Corlay, S.; et al. Jupyter Notebooks—A Publishing Format for Reproducible Computational Workflows. In *Positioning and Power in Academic Publishing: Players, Agents and Agendas*; Loizides, F., Schmidt, B., Eds.; IOS Press: Amsterdam, The Netherlands, 2016; pp. 87–90.
51. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
52. Buitinck, L.; Louppe, G.; Blondel, M.; Pedregosa, F.; Mueller, A.; Grisel, O.; Niculae, V.; Prettenhofer, P.; Gramfort, A.; Grobler, J.; et al. API Design for Machine Learning Software: Experiences from the Scikit-Learn Project. In Proceedings of the ECML PKDD Workshop: Languages for Data Mining and Machine Learning, Prague, Czech Republic, 23–27 September 2013; pp. 108–122.
53. Keras. 2015. Available online: <https://keras.io> (accessed on 1 November 2024).
54. Abadi, M.; Agarwal, A.; Barham, P.; Brevdo, E.; Chen, Z.; Citro, C.; Corrado, G.S.; Davis, A.; Dean, J.; Devin, M.; et al. Tensorflow: Large-scale machine learning on heterogeneous distributed systems. *arXiv* **2016**, arXiv:1603.04467.
55. Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; et al. PyTorch: An Imperative Style, High-Performance Deep Learning Library. In *Advances in Neural Information Processing Systems 32*; Curran Associates, Inc.: Red Hook, NY, USA, 2019; pp. 8024–8035.
56. Seabold, S.; Perktold, J. Statsmodels: Econometric and Statistical Modeling with Python. In Proceedings of the 9th Python in Science Conference (SciPy 2010), Austin, TX, USA, 28 June–3 July 2010; pp. 92–96.
57. Wiese, F.; Schlecht, I.; Bunke, W.D.; Gerbaulet, C.; Hirth, L.; Jahn, M.; Kunz, F.; Lorenz, C.; Mühlenpfordt, J.; Reimann, J.; et al. Open Power System Data—Frictionless data for electricity system modelling. *Appl. Energy* **2019**, *236*, 401–409. [CrossRef]
58. Nwaigwe, K.N. Assessment of wind energy technology adoption, application and utilization: A critical review. *Int. J. Environ. Sci. Technol.* **2022**, *19*, 4525–4536. [CrossRef]
59. Asiaban, S.; Kayedpour, N.; Samani, A.E.; Bozalakov, D.; De Kooning, J.D.M.; Crevecoeur, G.; Vandeveld, L. Wind and solar intermittency and the associated integration challenges: A comprehensive review including the status in the Belgian power system. *Energies* **2021**, *14*, 2630. [CrossRef]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.