*Accepted by Energy Conversion and Management. Formal publication version:* https://doi.org/10.1016/j.enconman.2022.116049

# Hourly solar irradiance forecasting based on encoder-decoder model using series decomposition and dynamic error compensation

Junlong Tong, Liping Xie*, Shixiong Fang, Wankou Yang and Kanjian Zhang

*School of Automation, Southeast University, Nanjing 210096, China*
*Key Laboratory of Measurement and Control of Complex Systems of Engineering, Ministry of Education, Nanjing 210096, China*

ARTICLE INFO

*Keywords*:
Solar irradiance forecasting
Deep learning
Temporal convolutional network
Encoder-decoder
Long short term memory
Error compensation

ABSTRACT

Accurate solar irradiance prediction is crucial for harnessing solar energy resources. However, the pattern of irradiance sequence is intricate due to its nonlinear and non-stationary characteristics. In this paper, a deep hybrid model based on encoder-decoder is proposed to cope with the complex pattern for hourly irradiance forecasting. The hybrid deep model integrates complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN), encoder-decoder module, and dynamic error compensation (DEC) architecture. The CEEMDAN is implemented to reduce the nonlinear and non-stationarity of the irradiance sequence. The encoder-decoder integrates temporal convolutional networks (TCN), long short-term memory networks (LSTM), and multi-layer perceptron (MLP) for temporal features extraction and multi-step prediction. The DEC architecture dynamically updates the model based on adjacent error information to mine the predictable components of error information. Furthermore, a new loss function is further proposed for multi-objective optimization to balance the performance of multi-step forecasting. In the hourly irradiance forecasting experiments on the three public datasets, the root mean square error (RMSE), mean absolute error (MAE), and correlation coefficient (R) of the proposed model are observed to be in a range of 30.693-34.433 W/m$^2$, 19.398-22.900 W/m$^2$, and 0.9872-0.9902, respectively. Compared to the benchmark models (including MLP, LSTM, and TCN), the RMSE and MAE reduce by 10.76%-22.00% and 5.47%-20.40%, respectively. The experimental results indicate that the proposed model shows accurate and robust forecasting performance and is a reliable alternative to hourly irradiance forecasting.

## 1. Introduction

Energy is the cornerstone of modern industry, which is of critical importance to the development of economics and life. Contrasting to the high rate of fossil fuel consumption, the replenishment is slow which has led to a significant increase in the energy crisis. Meanwhile, conventional energy sources are considered to be contributing significantly to greenhouse gas emissions and air pollution, which aggravates the greenhouse effect and environmental pollution.

Renewable energy sources are almost inexhaustible and cause little environmental or climate damage, which has captured wide attention around the world. In addition, solar energy is regarded as the most promising renewable energy source owing to its large amount and easy accessibility[1]. The average solar irradiance received by the surface of the earth is 1367W/m$^2$ about 1.8×10$^8$GW full year[2]. Such a huge form of energy has the potential to satisfy all the global demand for electricity. According to a report by REN21, total renewable energy generation in 2020 reaches 2838GW across the globe, of which solar power reaches 760GW, accounting for 26.77 percent[3].

Despite the high share of solar power generation in renewable energy generation, there are still many unresolved issues in the solar energy utilization process. Compared to conventional energy sources, the output of solar photovoltaic is quite unstable owing to the uncertainty and intermittency of solar irradiance and meteorological condition, which is seriously jeopardizing the stability and security of the grid[4]. Solar irradiance is the most influential factor in Photovoltaics(PV) power generation, indicating that accurate forecasting of solar irradiance variation is vital to PV power forecasting, PV grid integration, supply balance, and real-time dispatch[5].

The mainstream methods of solar irradiance forecasting can be classified into three categories[6, 7, 8], i.e., physical models, classical statistical models, and machine learning models. Physical models mainly refer to utilizing numerical weather prediction(NWP) to forecast solar irradiance[9]. The method employs the conditions of atmospheric, oceanic motion and current weather information to build a mathematical model. This mathematical model serves to establish the relationship between the solar irradiance and the physical condition such as atmospheric related physical factors. Physical models can produce large-scale and long-term predictions, which are suitable for long-term forecasting task. On the contrary, due to its excessive complexity, remarkably consuming of the computational resources, and the low output frequency of mainstream NWP systems, the NWP model may cause significant delays when used for short-term predictions[10]. Statistical models forecast the future solar irradiance depending on historical trends of solar irradiance. The classical statistical models include autoregressive moving average[11], multiple linear regression[12], and autoregressive integrated moving average[13]. But the simple structure of these models makes it difficult to fit the fluctuation of irradiance caused by the changes of external environment in a short-term. Besides, machine learning methods have created a sensation in areas such as computer vision[14] and expression detection[15]. In recent years, machine learning methods have been widely used for irradiance prediction, such as support vector machines (SVM)[16], random forests[17], and neural networks [18]. These models are data-driven, which can learn the intrinsic connections between input and output from a large amount of data[19]. Furthermore, the machine learning models can effectively extract temporal features of irradiance itself and establish the mapping relationship between irradiance and covariables (such as environmental factors). Besides, deep learning approaches have begun to receive the attention of researchers in the area of time series forecasting. Literature[20] reviews 232 papers on irradiance prediction based on machine learning methods over the last 20 years and discusses the existing machine learning models. The results show that the concept of feature extraction based on deep learning models has been prevalent in existing solar radiation prediction studies.

Nonetheless, owing to the superposition of different meteorological factors, there is pattern mixing in irradiance sequences, i.e., the sequences contain features of multiple patterns, and such mixed patterns make the task of feature extraction difficult. In addition, the

---

*Corresponding author
Email address: lpxie@seu.edu.cn

## Nomenclature

| | |
|---|---|
| ARIMA | Auto regressive integrated moving average |
| Bi-LSTM | Bidirectional long short-term memory |
| CEEMDAN | Complete ensemble empirical mode decomposition with adaptive noise |
| CNN | Convolutional neural network |
| CWT | Continuous wavelet transform |
| DEC | Dynamic error compensation |
| EC | Error compensation |
| EEMD | Ensemble empirical mode decomposition |
| ELM | Extreme learning machine |
| EMD | Empirical mode decomposition |
| IMF | Intrinsic mode function |
| LSTM | Long short-term memory |
| MAE | Mean absolute error |
| MLP | Multilayer perceptron |
| nRMSE | Normalized root mean square error |
| NWP | Numerical weather prediction |
| PV | Photovoltaics |
| R | Pearson correlation coefficient |
| ReLU | Rectified linear unit |
| RMSE | Root mean square error |
| RNN | Recurrent neural network |
| SVM | Support vector machine |
| TCN | Temporal convolutional network |
| VMD | Variational mode decomposition |
| WPD | Wavelet packet decomposition |

distinct periodic characteristics of irradiance such as different seasons corresponding to different features require prediction models to own the ability to maintain long-term dependencies. It is difficult for a single deep prediction model to solve the above problems. Therefore, this work proposes a hybrid deep learning model to try to challenge these difficulties. Adopting the hybrid models has been recognized as an effective way to deal with the strong instability and fluctuation of irradiance by incorporating the advantages of different models[21].

The solar irradiance is influenced by many meteorological factors such as wind speed, air pressure, temperature and humidity, some of which may have a coupling relationship with other factors, which brings about complex nonlinear and non-stationary characteristics. Signal decomposition is considered an effective method to reduce the non-linearity and non-stationarity of time series[22], which are widely used in solar irradiance forecasting. The mainstream decomposition methods include empirical mode decomposition (EMD), ensemble empirical mode decomposition (EEMD), wavelet decomposition (WD), and wavelet package decomposition (WPD). Lan et al.[23] preprocessed solar irradiance data using the EEMD algorithm and utilized a self-organizing map and back-propagation neural network model for day-ahead irradiance prediction, where the EEMD is used to decompose the irradiance into several subsequences as additional features. Experimental results show that the decomposition method can improve the prediction performance. Combining the SVM and WD algorithm to forecast solar irradiance is proposed by Mohammadi et al.[24], which SVM is used to forecast the subseries decomposed by WD. A hybrid model based on a multi-branch hybrid structure composed of 1D-convolution and long short-term memory (LSTM) was used for irradiance prediction, which uses WPD to decompose the irradiance and predict the subsequences separately[21]. Monjoly et al.[25] forecasted solar irradiance by combining three multiscale decomposition methods (including EMD, EEMD, and WD) with artificial neural networks (ANN) models, where the EMD, EEMD, and WD are used to decompose the clear-sky index. The experimental results show that the forecasting performance is improved by introducing signal decomposition.

The temporal features extraction is a key step in the irradiance forecasting task, which is usually performed by deep learning models. Several studies focused on the recurrent neural network (RNN) and convolutional neural network (CNN), which validated the effectiveness in the field of solar irradiance prediction from different perspectives. A study utilizing the LSTM model for hourly solar irradiance forecasting proposed by Qing et al.[26] shows its superiority when compared with multilayer perceptron (MLP). Acikgoz[27] proposed a method that uses continuous wavelet transform (CWT) to convert the time-series data into image data to facilitate CNN structure for feature extraction such as GoogleNet and Alexnet, and finally to predict irradiance using an extreme learning machine, shown a great forecasting performance. Some researchers attend to the study about the deep learning forecasting methods from the perspective of hybrid structures with CNN and RNN. Zang et al.[28] proposed a method for combining meteorological information of adjacent sites, which uses a CNN-LSTM cascade structure as a feature extractor to extract spatial features of adjacent sites, as well as temporal features of the respective sites. Kumari et al.[29] analysed the impact yielded by the order of CNN and LSTM modules on system performance. Gao et al.[30] proposed a variety of CNN-LSTM-based structures for solar irradiance forecasting, where they compared the performance of different combination structures and obtained an optimal combination structure. The experimental results of these hybrid structures illustrate that the strategy can effectively improve the forecasting performance.

In addition, most of the existing time series forecasting models concentrate only on the preliminary forecasting results and neglect to process the forecasting errors[31]. The prediction errors of several adjacent time steps may be correlated, meaning that there are still predictable components in the error series. Depending on the experimental results of related researches, correcting errors may enhance the forecasting performance[32]. Ma et al.[33] performed an error compensation strategy for wind speed forecasting, where the LSTM model is employed to forecast error series. Four real cases demonstrate the effectiveness of the error compensation strategy by using LSTM. Sun et al.[34] presented a wind power forecasting method, where random forest algorithm, k-means clustering and LSTM are employed to infer the latent characteristics of preliminary series. Besides, LSTM is used to model the error sequences to mine the inherent information. The validity of the proposed error compensation model is illustrated by experiments from two wind farms. Ma et al.[35] proposed a deep learning method for PV power forecasting, where a model based on extreme learning machine (ELM) algorithm is established to forecast error series. In addition, several works construct error forecasting models based on error decomposition. Liu et al.[31] proposed an error compensation method for wind speed forecasting, where an ELM model is employed to forecast the error sub-sequences decomposed by empirical wavelet transform.

### 1.1. Limitations of the above literature

(1) Several decomposition methods mentioned in the literature review have been widely used. However, an essential defect of the EMD algorithm is the phenomenon of mode mixture, namely a single intrinsic mode function (IMF) consisting of variously scaled signals. The EEMD method overcomes the mode mixing problem, but the residual white noise is not negligible after reconstruction, which brings difficulties to prediction. Besides, the decomposition results of WD and WPD are determined by the wavelet basis function[36], which is generally obtained by the empirical selection, which introduces the human factor into the decomposition result. (2) The hybrid structure is regarded as combining the ability to extract spatial features of CNN with the ability to extract temporal features of RNN[37]. However, typical CNN architectures extract features by sliding convolutional kernels in space and lack causal mechanisms in the movement process, i.e., the values of future moments are used to compute the feature representation of the current moment. In essence, typical CNN architecture has insufficient ability to capture sequential features. Besides, the scalability of the above hybrid model for multi-step prediction tasks is limited, since different models have to be trained for different steps in prediction. (3) Although some residual correction models for prediction error sequences have been proposed in the above literature, they are essentially two-stage models, i.e., two models are used to predict the preliminary series and error series separately. The two-stage models introduce the complexity in adjusting the parameters and the error information cannot be fed directly to the prediction mode. If only one end-to-end model is utilized to perform both the preliminary prediction and the error prediction, it will reduce the amount of the model hyperparameters and updated the model based on the error information dynamically.

### 1.2. Contributions and innovations

According to the above analysis, a hybrid model based on deep learning methods is proposed to perform hourly solar irradiance forecasting, which is composed of reduction of signal non-stationarity, feature extraction, multi-objective optimization, and dynamic error compensation. Specifically speaking, the proposed hybrid method consists of three parts: series decomposition module, encoder-decoder module, and error compensation module. Firstly, the complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN) is implemented to reduce the nonlinear and non-stationarity of the irradiance series, where the CEEMDAN is an adaptive decomposition method without providing basis function. Besides, the method eliminates auxiliary white noise in ensemble averaging to overcome the disadvantages of large reconstruction error and noise cancellation in EEMD. The adoption of the CEEMDAN algorithm can automatically subdivide the original nonstationary irradiance sequence into several sub-sequences of different modes with better stability.

Then each sub-sequence is encoded as a semantic vector by the encoder module, which is the input of the decoder module. The proposed encoder-decoder is a sequence-to-sequence model with better scalability for multi-step forecasting, which can be achieved by an autoregressive method. The encoder uses a cascaded structure of temporal convolutional network (TCN) and LSTM, which combines the advantages of the large receptive field and multilayer stacking of TCN, and the advantages of long-term dependencies of LSTM. Furthermore, due to the causal mechanism, the TCN is more suitable for temporal feature extraction compared with typical CNN structure. The decoder that consists of a parallel connection of LSTM-MLP module can output the results of multiple prediction steps by means of autoregression. Meanwhile, to balance the performance of multi-step forecasting, a new loss function is designed to implement a multi-step constraint.

In the error compensation module, the error information is added to update the model parameters. Different from the two-stage methods, the proposed dynamic error compensation (DEC) method uses a single end-to-end model to perform both the preliminary prediction and error prediction. The DEC technique dynamically updates the model parameters by forming a sequence of forecasting errors for several moments before the current moment. Then the error sequence will be a part of the input features, which can compensate for the deficiency of the model for the current input series. This technique can effectively reduce the forecasting error and improve the stability and robustness of the model only using one end-to-end model.

Several experiments are designed in different datasets to verify the validity of the proposed model.

The contributions of the work can be summarized as follows:

1. **An adaptive decomposition algorithm is conducted to reduce the nonlinearity and instability of the solar irradiance series.**
2. **A novel encoder-decoder architecture based on the deep learning methods is designed for feature extraction and representation of the relationship between solar irradiance and its corresponding meteorological information.**
3. **Multi-objective optimization method is employed to balance the performance of multi-step outputs.**
4. **A dynamic error compensation strategy is proposed to process the residual error information.**

The remainder of the paper is arranged as follows: The detail of the architecture of the proposed model is presented in Section 2; Section 3 carries out several experiments to test the performance of the proposed model; The conclusions of the study are shown in Section 4.

## 2. Methodology

### 2.1. Framework of the proposed method

An overview of the proposed forecasting method is presented in Fig. 1. The proposed method can be divided into three parts: (1) Time series decomposition module, which is applied to decompose the irradiance sequence into several sub-sequences to increase the stationary characteristics of the model input signal. (2) Encoder-decoder architecture, which is adapted to extract temporal features and forecast solar irradiance. (3) Dynamic error compensation strategy, which is designed to update the model based on error information in an adaptive way.

### 2.2. Time series decomposition module

The complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN)[38] is an improved variant of the EMD method that provides a precise reconstruction of the original time series, a lesser number of iterations, and a better spectral separation of the IMFs by adding a specific noise at decomposition step and calculating a particular residue to get each mode.

Let $x(t)$ be the input signal and $w^i(t)$ be white Gaussian noise with $\mathcal{N}(0,1)$. Assign $IMF_k(t)$ as the $k-th$ mode of input signal. The CEEMDAN algorithm can be described as follow:

1. The $x^i(t)$ are generated through the addition of various realizations of white noise $w^i(t)$ to the input signal $x(t)$, which can be expressed in eq. (1). Assigning $IMF_k^i$ as the $k-th$ mode of $x^i(t)$ decomposed by EMD, the first mode of the input signal is calculated by eq. (2):

$$x^i(t) = x(t) + \varepsilon w^i(t), i = 1, 2, ..., I, \tag{1}$$

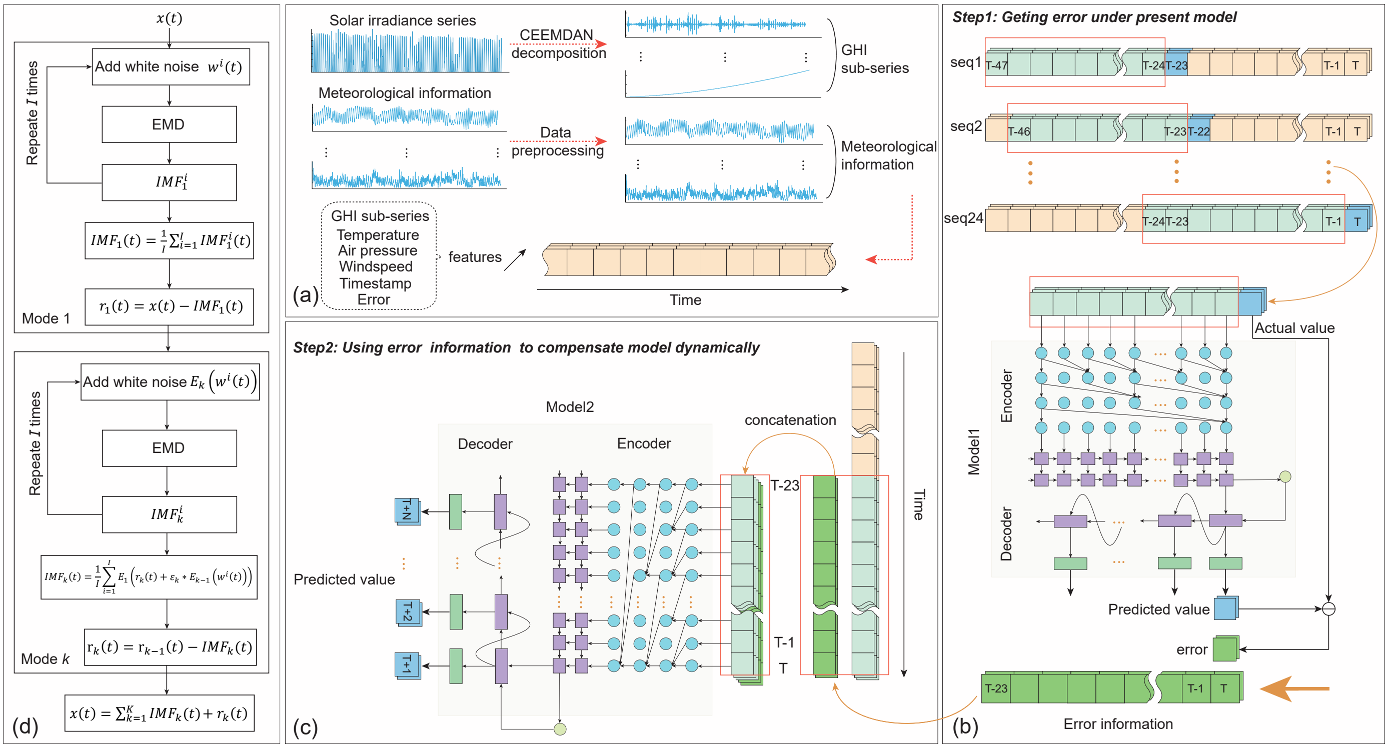

**Figure 1:** Framework of the proposed forecasting method. (a) Time series decomposition and features arrange; (b) Process of getting error under present model; (c) Process of updating temporal feature extraction model with error compensation; (d) Flowchart of the signal decomposition module.

$$IMF_k(t) = \frac{1}{I}\sum_{i=1}^{I} IMF_k^i(t), \qquad (2)$$

where $I$ represents the number of realizations and $\varepsilon$ is noise standard deviation. Both of $I$ and $\varepsilon$ are hyperparameters.

(2) Calculate the first residual at the stage k=1 : $r_1(t) = x(t) - IMF_1(t), i = 1, 2, ..., I$. Then the signal $r_1(t) + \varepsilon E_k(w^i(t))$ can be decomposed by EMD to obtain $IMF_2(t)$ seen as eq. (3):

$$IMF_2(t) = \frac{1}{I}\sum_{i=1}^{I} E_1\left(r_1(t) + \varepsilon_1 E_k\left(w^i(t)\right)\right), \qquad (3)$$

where $E_k(\cdot)$ is defined as an operation to generate $k-th$ mode that decomposed by EMD.

(3) For $k = 2, ..., K$, the $k-th$ residual $r_k(t)$ is calculated by eq. (4):

$$r_k(t) = r_{k-1}(t) - IMF_k(t). \qquad (4)$$

(4) The $IMF_{k+1}(t)$ is calculated by eq. (5):

$$IMF_{k+1}(t) = \frac{1}{I}\sum_{i=1}^{I} E_1\left(r_k(t) + \varepsilon_k E_k\left(w^i(t)\right)\right). \qquad (5)$$

(5) Repeat steps 3 and 4 for the next $k$ until cannot find a new mode, i.e. the residual $r_K(t)$ is no longer suitable for decomposition. Hence the number of IMF is determined by the algorithm itself.

The residual $R(t)$ of the CEEMDAN can be represented by eq. (6):

$$R(t) = x(t) - \sum_{k=1}^{K} IMF_k(t). \qquad (6)$$

Besides, the flowchart of the CEEMDAN algorithm is shown in Fig. 1(d) to explain the decomposition process.

### 2.3. Encoder-decoder architecture

One of the difficulties in irradiance forecasting is how to accurately model the impact of weather information and historical data on solar irradiance. The encoder-decoder architecture based on deep learning methods can extract the temporal features and learn the internal correlation of temporal features efficiently.

Encoder-decoder is a framework for dealing with sequence-to-sequence problems, which was first applied in the field of machine translation[39]. The major advantage of the framework is the capacity to process lateral length data. For the irradiance prediction problem, the structure of the decoder can be designed to obtain the output results of multiple time steps at once. In this work, we proposed a novel encoder-decoder architecture to extract temporal features and forecast irradiance, which is shown in Fig. 2(a). Different from the traditional RNN encoder, the significant features of our method reflect in using the cascaded network of TCN and LSTM as the encoder and using LSTM and multi-layer perception as decoder.

The significant advantages of the TCN structure are the large receptive field as well as the ability to stack multiple layers, which means that the encoder can achieve deeper networks and accept more historical information. The LSTM structure is capable of learning order dependence in sequence forecasting problems, which means that the feature of long-term dependence on sequence interrelations is one of the advantages of the encoder structure. By utilizing the cascaded TCN and LSTM modules, we find that the encoder can obtain better performance for temporal feature extraction.

The decoder is composed of LTSM module and MLP module, where the LSTM module is applied to decode the state vector and the function of the MLP is dimension transformation to match the dimension of desire output. Furthermore, the forecasting results of different time steps can be inferred by a multi-step output decoder, where the decoder stacks

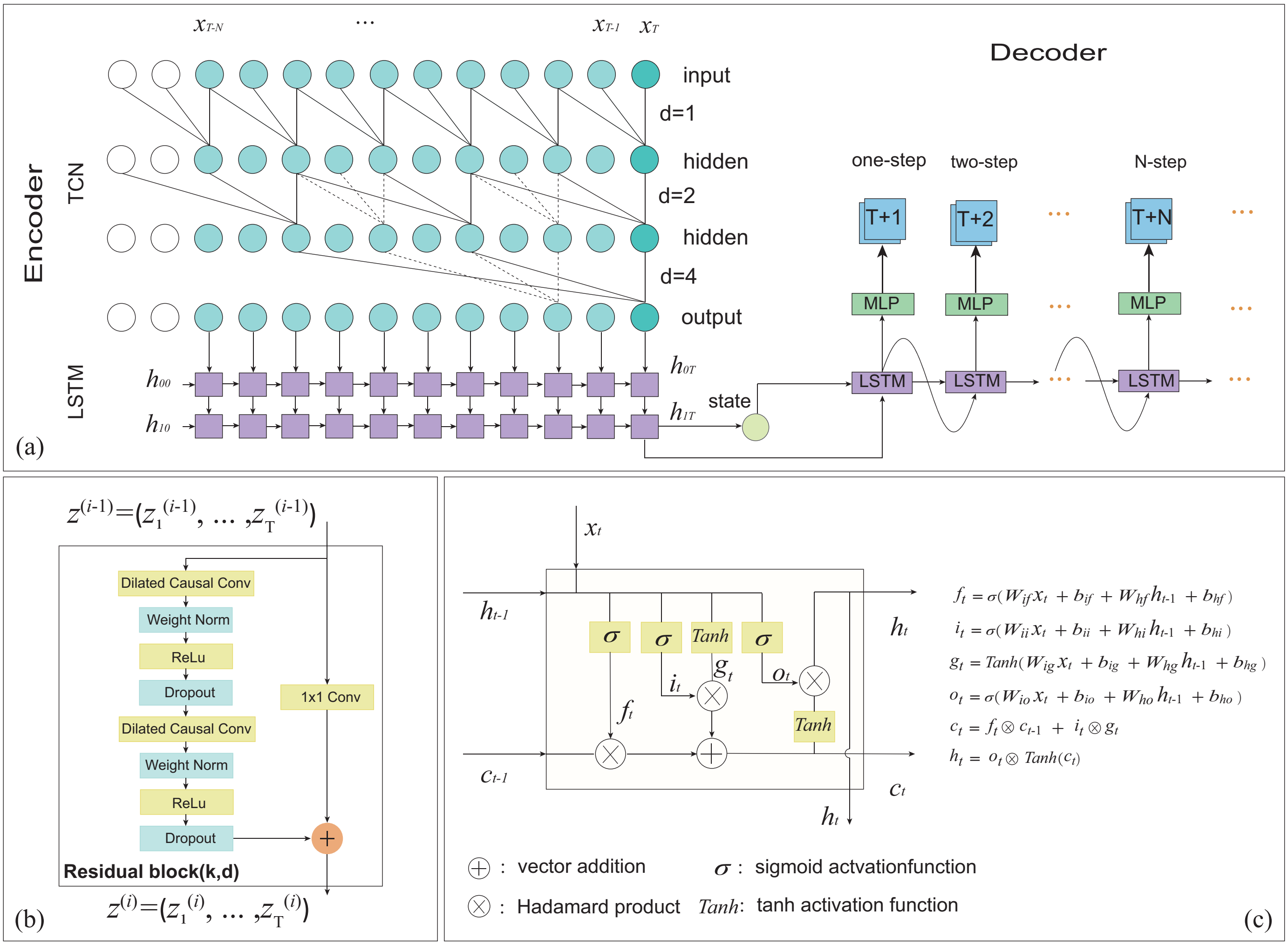

**Figure 2:** Elements in the proposed encoder-decoder architectural. (a) Framework of the encoder-decoder, which a dilated causal convolution with dilation factors $d = 1, 2, 4$ and kernel size $k = 3$; (b) TCN residual block. Adding a $1 \times 1$ convolution to block for dimensional transformation; (c) Internal structure of a LSTM module.

the assembly unit, namely LSTM modules and MLP units, along the temporal direction.

To improve the generalization performance[40, 41], a new loss function is designed for the constraints of the multi-step outputs model. The performance of different time steps can be balanced by superimposing the losses of different time steps. By applying different weights to the decoder outputs, different degrees of emphasis on the outputs of different time steps can be achieved. The loss function can be shown in eq. (7):

$$Loss = \sum_{i=1}^{K} \alpha_i \cdot loss_i + \beta \cdot \|\omega\|^2, \tag{7}$$

where $K$ is the length of the output, $\alpha_i$ represents the weight of loss in step $i$, $\omega$ means the parameter of the model, and $\beta$ is the regularization coefficient of the L2 regularization.

Furthermore, the details of the components of the proposed encoder-decoder are as follows:

**TCN:** Temporal convolutional network (TCN)[42] was designed as a generic convolutional structure for temporal feature extraction and sequence modeling, which the structure has two notable features that causal constraint and equal length mapping. For a task of sequence modeling, suppose that $y_0, y_1, ..., y_t$ is the output sequence and the corresponding sequence $x_0, x_1, ..., x_t$ is assigned as the input sequence. The casual constraint means that there is no information leakage, in other words, $y_t$ hinges only on the historical sequence $x_0, x_1, ..., x_t$. Equal length mapping refers to the length of the output sequence being the same as the input sequence in a TCN model, which is designed to facilitate layer stacking.

In order to follow the two principles and achieve a long history size, causal convolutions based on 1-D fully-connected network (FCN) and dilated convolutions are employed to the TCN structure. The 1-D-FCN-based causal convolutions obtain the output at time $t$ by convolving the elements at time $t$ and before in the previous layer along the time direction. The dilated convolutions exponentially expand the cavity spacing as the number of layers increases to increase the perceptual field, which allows for more historical data to be considered. Specifically, for a kernel $f : 0, 1, ..., k-1 \rightarrow \mathbb{R}$, the dilated convolution $F$ based on $f$ on element $s$ of the 1-D input sequence $\mathcal{X}$ satisfies eq. (8):

$$F(s) = (\mathcal{X} *_d f)(s) = \sum_{i=0}^{k-1} f(i) \cdot \mathcal{X}_{s-d \cdot i}, \tag{8}$$

where $*$ means convolutional operation, $d$ is the dilation factor, $k$ accounts for the kernel size, and $s - d \cdot i$ represents the direction of the past. If $d = 1$, the dilated convolution is equivalent to 1-D FCN. An example of TCN structure can be shown as Fig. 2(a) with dilation factors $d = 1, 2, 4$ and kernel size $k = 3$ and 2 hidden layers.

Moreover, a residual connection is adopted to the TCN model for a deeper and larger structure. The residual block of TCN is comprised of the residual connection and two layers of dilated causal convolution, as presented in Fig. 2(b). After each dilated causal convolution layer, the rectified linear unit (ReLU) is used as the nonlinear units. In addition, the batch normalization and the dropout layer are adopted.

**LSTM:** Long short term memory network (LSTM)[43] was proposed to address the defect of long-term memory in recurrent neural networks, in which the most important feature is the introduction of three gate structures, i.e., forget gate, input gate, and output gate. By processing the information flow with the three gate structures, LSTM can achieve long-term dependence and overcome the gradient vanishing and gradient exploding problems. The internal structure and the general data flow calculation process of LSTM are shown in Fig. 2(c), where $W, b$ are the parameters of LSTM and $h, c$ are the hidden states of LSTM. Besides, $f_t$, $i_t$ and $o_t$ represent the output of forget gate, input gate, and output gate respectively. The hidden states are used as the input of the LSTM to ensure that the historical information can be retained. The detailed calculation process of LSTM can be found in the literature[44].

The staking method of LSTM can be depited in Fig. 2(a), which presents a two-layer LSTM network in the encoder and a single layer LSTM network in the decoder.

### 2.4. Dynamic error compensation strategy

Due to the irradiance prediction problem with nonlinear and non-stationary characteristics, it is inevitable that a time series forecasting model produces errors. Compensating the prediction model with error information is a solution that can be considered. We design an error compensation structure based on the assumption that the prediction errors of several adjacent time steps are correlated[31].

Different from two-stage error compensation methods, which simply add the irradiance forecasting value and the error forecasting value at moment $T + 1$, the proposed strategy does not forecast the error at moment $T + 1$ but obtains the error at each moment of the input sequence with the irradiance forecasting model. The parameters of the irradiance forecasting model can be dynamically updated based on the error information corresponding to the input sequence, where the error sequence is generated by utilizing the current model to forecast irradiance for several moments before the moment $T + 1$. Then the updated model is used to forecast the irradiance at the next moment. The strategy is also suitable for other temporal feature extraction models.

The major process of dynamic error compensation strategy can be described in Fig. 1. In order to illustrate the process more conveniently, we stipulate that the predicted moment is $T + 1$ and the sequence information of the previous 24 hours, i.e. from $T - 23$ to $T$, is used as input information to forecast irradiance at moment $T + 1$.

Firstly, the sequence of length 48 before moment $T + 1$ is divided into 24 subsequences through the sliding window method, where the length of the sliding window set as 24 and the sliding step set as 1. As shown in Fig. 1(b), the irradiance forecasting error at moment $T - 23$ can be obtained by comparing the actual value with the forecast value of *model1*, which utilizes sub-sequence from moment $T - 47$ to $T - 24$ as input. Analogously, based on sequentially moving the sliding window and inputting subsequence to *model1*, the irradiance forecasting error at each other moment can be obtained.

As presented in Fig. 1(c), the error sequence is concatenated with the original feature sequence along the feature direction as a new feature to input model, which the error sequence consists of the forecasting errors of the current model parameters at each moment of the input time series. The model parameters can be dynamically updated to reduce the irradiance forecasting error at moment $T + 1$ according to the error information of the past 24 hours.

Note that in order to implement the dynamic error compensation strategy, $model1$ and $model2$ are restricted to the same structure and parameters, which requires the input dimension of two models to be consistent. Zero-padding the input sequence feature of $model1$ is a viable solution to ensure the same input dimension of the two models. During the training stage, *model1* is only used to obtain the irradiance forecasting error of the moment from $T - 23$ to $T$ without updating parameters, whereas *model2* updates the parameters according to the original sequence and error sequence from moment $T - 23$ to $T$.

## 3. Experiment and analysis

### 3.1. Dataset description

In this work, three sites with different climatic conditions were selected for model performance evaluation, where the there publicly available datasets consisting of solar irradiance and corresponding meteorological information (temperature, humidity, wind speed, etc.). The first two datasets from Hawaii (19°73′N 156°06′W, USA) and Oak Ridge (35°93′N 84°31′W, USA) are composed of 5 years of hourly solar irradiance data and historical meteorological data measured from 1st January 2013 to December 31, 2017, which are provided by the National Renewable Energy Laboratory (NREL)[45]. The last datasets were collected from DKA Solar Centre (DKASC)[46], which is located at the Desert Knowledge Precinct in Alice Springs of Australia. The Alice Spring dataset of DKASC used in this work is ranging from 1st January 2010 to December 31, 2014, which consists of hourly solar irradiance data and historical meteorological data. The detailed information of the two datasets are provided by Table. 1.

Each dataset is divided into the training set (containing approximately 35,040 samples over 4 years) and the testing set (containing approximately 8,760 samples over 1 year), respectively. The detail information of the datasets can be found in Table. 2.

### 3.2. Data processing

Since irradiance variation is directly influenced by external meteorological conditions, historical meteorological data such as wind speed, humidity, pressure and temperature are used as influencing factors in this study. In addition, temporal information can effectively characterize diurnal and seasonal effects on irradiance, such that temporal information should be used as an important feature of model input.

In this work, a cyclical features encoding method is applied to encode temporal information, which adopts both sine and cosine functions to normalize the temporal information and achieve effective differentiation of the temporal information while preserving periodicity. In comparison to other encoding methods such as one-hot encoding, the cyclical features encoding enables the efficient preservation of temporal information with periodic characteristics. The encoding formulas can be shown in eq. (9):

$$\begin{aligned} \mathrm{Tsin} &= \sin(2\pi \frac{t}{T}), \\ \mathrm{Tcos} &= \cos(2\pi \frac{t}{T}), \end{aligned} \tag{9}$$

where $t$ is the current time and $T$ is cycle length.

Besides, it is necessary to perform operations such as missing data completion and normalization of the original data. For each attribute of a sample, the normalization process is performed as shown in eq. (10):

$$a' = \frac{a - a_{min}}{a_{max} - a_{min}}, \tag{10}$$

**Table 1**
General information about the three datasets.

| Dataset | Location | Climate type | Meteorological information | Minimum | Maximum | Mean | Standard deviation |
|---|---|---|---|---|---|---|---|
| Hawaii dataset | 19°73′ N 156°06′ W | Rainforest | GHI(W/m$^2$) | -4.47 | 1100.57 | 239.91 | 324.84 |
| | | | temperature(°C) | -9.85 | 37.01 | 26.31 | 3.16 |
| | | | humidity(%) | 26.69 | 97.14 | 66.27 | 9.49 |
| | | | wind speed(m/s) | 0.02 | 12.59 | 2.52 | 1.32 |
| Oak Ridge dataset | 35°93′N 84°31′W | Humid subtropical | GHI(W/m$^2$) | -0.01 | 1003.68 | 158.82 | 239.61 |
| | | | temperature(°C) | -19.51 | 37.69 | 15.52 | 10.04 |
| | | | humidity(%) | 19.19 | 98.96 | 77.22 | 17.90 |
| | | | wind speed(m/s) | 0.20 | 5.19 | 0.83 | 0.70 |
| Alice Spring dataset | 23°41′S 133°52′E | Desert | GHI(W/m$^2$) | 1.14 | 1688.19 | 253.72 | 351.21 |
| | | | temperature(°C) | -5.60 | 44.81 | 20.80 | 9.53 |
| | | | humidity(%) | 3.33 | 101.79 | 42.52 | 25.59 |
| | | | wind speed(m/s) | 0 | 20.01 | 2.55 | 1.68 |

**Table 2**
Information of dataset segmentation.

| Dataset | Period(year) | Training set(year) | Testing set(year) |
|---|---|---|---|
| Hawaii dataset | 2013-2017 | 2013-2016 | 2017 |
| Oak Ridge dataset | 2013-2017 | 2013-2016 | 2017 |
| Alice Spring dataset | 2010-2014 | 2010-2013 | 2014 |

where $a$ is the actual value of an attribute of feature space, $a_{min}$ is the minimum of the attribute, and $a_{max}$ is the maximum of the attribute.

### 3.3. Evaluation metrics and experimental Setup

In order to provide a comprehensive assessment of the proposed models, four mainstream Evaluation metrics are used in this work, namely root mean squared error (RMSE), mean absolute error (MAE), normalized root mean squared error (nRMSE), and Pearson's correlation coefficient (R).

RMSE is used to assess the average squared difference between the predicted values and the actual value, which is used to compare forecasting errors of different models for a specific dataset. The model with RMSE metric will assign higher weights to the outliers, and the magnitude of the outliers can be measured by RMSE. The RMSE can be defined as eq. (11):

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}. \tag{11}$$

MAE measures the average of all absolute errors that is the absolute average distance between the actual data and the predicted data. MAE preserves the same units of measurement as the data under analysis and gives all individual errors the same weights, which can describe the prediction performance from an average perspective. The formula of MAE can be shown as eq. (12):

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}\left|y_i - \hat{y}_i\right|. \tag{12}$$

NRMSE is the normalized root mean square error that is used for calculating scale-independent errors. The normalized form converts the nRMSE to a percentage, rendering it scale-free. The scale-free form provides a relative evaluation metric to compare models with different scales. The formula of nRMSE can be shown as eq. (13):

$$\mathrm{nRMSE} = \frac{\mathrm{RMSE}}{\bar{y}}. \tag{13}$$

**Table 3**
Hyperparameters of CEEMDAN-Encoder-decoder-EC model for 3-step ahead forecasting task.

| Hyperparameter | Value |
|---|---|
| Noise standard deviation ($\varepsilon$) | 0.02 |
| Number of realizations ($I$) | 500 |
| Number of tcn layers | 6 |
| Activation function of tcn module | ReLU |
| Number of hidden layers of lstm module | 2 |
| Hidden size of lstm module | 128 |
| Number of MLP layers | 1 |
| Batch size | 16/32 |
| Learning rate | 0.001/0.002 |
| Learning rate decay | 0.1/10 epochs |
| Weight coefficient of the loss function | (0.5,0.5,0.5) |

In addition, eq. (14) is Pearson's correlation coefficient (R) that is a metric to evaluate the linear correlation between two sets of data[47]. For forecasting task, the RMSE and the MAE only describe the prediction bias and cannot measure correlation, while the correlation coefficient R is used to describe the strength of the relationship between predicted series and actual series.

$$\mathrm{R} = \frac{\sum_{i=1}^{N}(y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum_{i=1}^{N}(y_i - \bar{y})^2}\sqrt{\sum_{i=1}^{N}(\hat{y}_i - \bar{\hat{y}})^2}}, \tag{14}$$

where $y_i$ represents the actual value of solar irradiance, $\hat{y}_i$ equates the forecasting value of solar irradiance, $\bar{y}$ is the mean value of $y_i$, $\bar{\hat{y}}$ is the mean value of $\hat{y}_i$, and $N$ is the number of samples in the dataset.

The CEEMDAN-Encoder-decoder-EC model is comprised of the Encoder-decoder module, the decomposition module, and the dynamic error compensation module. The inputs to the CEEMDAN-Encoder-decoder-EC model include the historical irradiance sequence and the corresponding covariates (i.e., humidity, temperature, wind speed, and time code). The hyperparameters of the proposed CEEMDAN-Encoder-decoder-EC model for 3-step ahead forecasting task can be found in Table. 3.

The selection of hyperparameters follows some experiences and rules as follows. Firstly, from the perspective of model architecture, the hyperparameters include the number of TCN layers and the number of LSTM layers. The number of TCN layers is determined according to the length of the input sequence, where too many layers will lead to too much padding operations resulting in useless information while

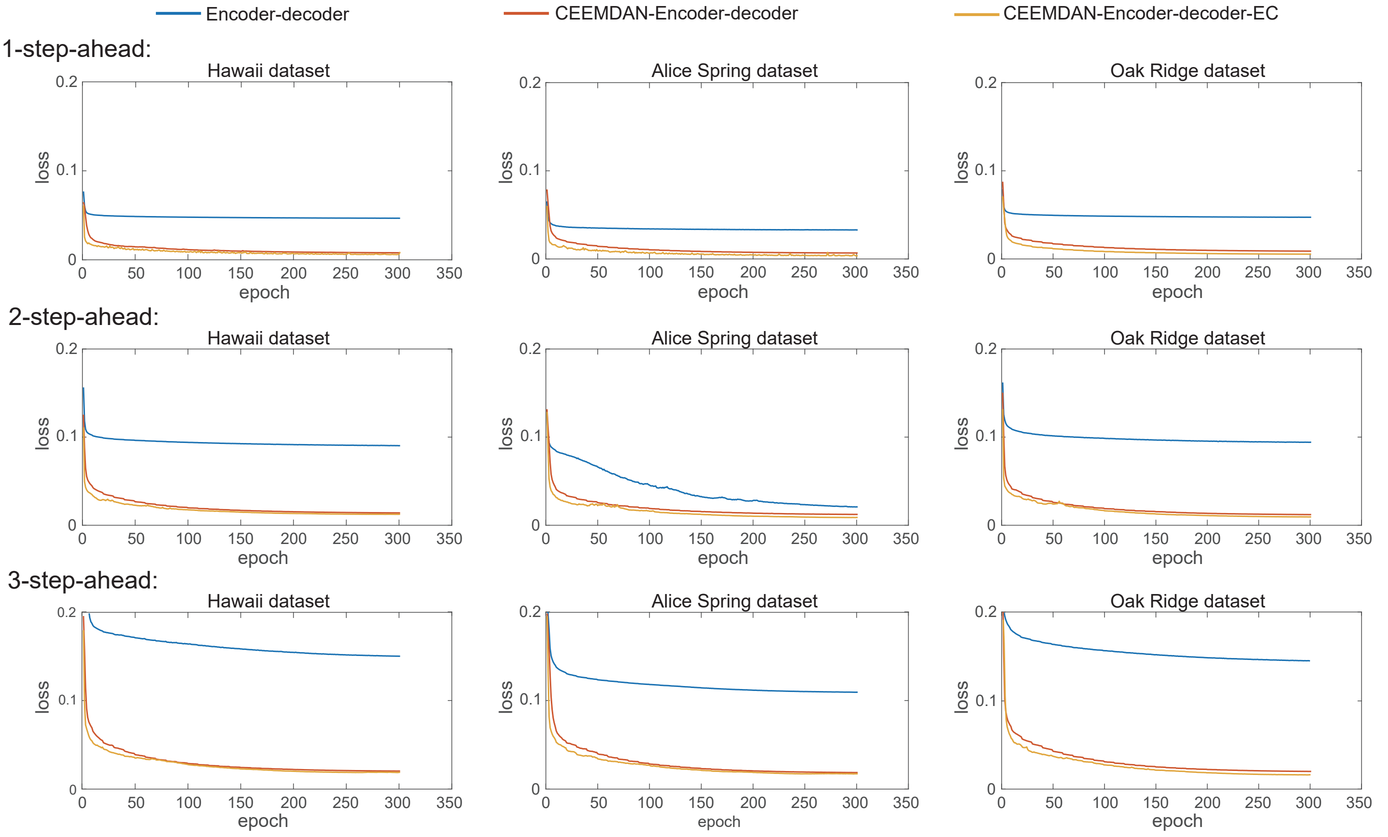

**Figure 3:** Training processes of proposed methods.

too few layers will lack the extraction of global information. Therefore, the criterion for determining the number of TCN layers is to utilize as much global information as possible while preserving as little padding operations as possible. For the LSTM, it is usually stacked 2 to 6 layers because too deep LSTM vertical stacking can lead to difficult training. In addition, the selection of learning rate and batch size follows the linear scaling rule[48], i.e., the learning rate and batch size should be changed proportionally when adjusting them. Grid search can be used to determine the appropriate hyperparameters. Besides, reinforcement learning[49] is an alternative to hyperparameter search. For a complex scenario such as inputting longer sequence, automated reinforcement learning (Auto-RL)[50] is an effective alternative due to the computational cost of grid search.

In this work, all the experiments are implemented with PyTorch 1.5 on TITAN RTX 24GB GPU under Ubuntu20.0 environment.

## 3.4. Analysis of the proposed method

The proposed model contains a series decomposition module to reduce nonlinear and non-stationary of the irradiance signal, the Encoder-decoder module to forecast the irradiance of the next moment, and the error compensation module to reduce prediction errors. In order to comprehensively verify the effectiveness of each module, we analyze the convergence and perform ablation experiments in this section.

### *3.4.1. Convergence analysis*

The training loss curves of the proposed models are shown in Fig. 3, which indicates the effect of each module on the convergence of the model. The three columns in Fig. 3 represent convergence over three different datasets, and each row represents convergence at a single stage.

It can be concluded from the Fig. 3 that:

(1) During the training process, the loss value of each model decreases quickly and eventually remains stable, which demonstrates that the proposed model shows well convergence performance.

(2) The proposed models exhibit well convergence on forecasting tasks in different datasets and with different time steps, which illustrates that the proposed model possesses great robustness.

(3) The decomposition module and the error compensation module can considerably accelerate the loss reduction of a model, which indicates that the two modules can improve the stability for the irradiance forecasting model.

### *3.4.2. Ablation study*

In order to evaluate the effect of different modules on the prediction model, ablation experiments with three comparison models are designed in this section. The Encoder-decoder module is used as a benchmark model. The CEEMDAN-Encoder-decoder model is composed of the Encoder-decoder module and the decomposition module. The CEEMDAN-Encoder-decoder-EC model is comprised of the Encoder-decoder module, the decomposition module, and the error compensation module. A series of experiments with different output steps are implemented on three different datasets.

Fig. 4 shows the scatter distribution of the actual values and predicted values on the Hawaii dataset, where the sparsity of the scatter diagram corresponds to the performance of the forecasting model. The first column represents the scatter diagram of the Encoder-decoder model, the second column is the scatter diagram of the CEEMDAN-Encoder-decoder model, and the last column demonstrates the performance of the CEEMDAN-Encoder-decoder model.

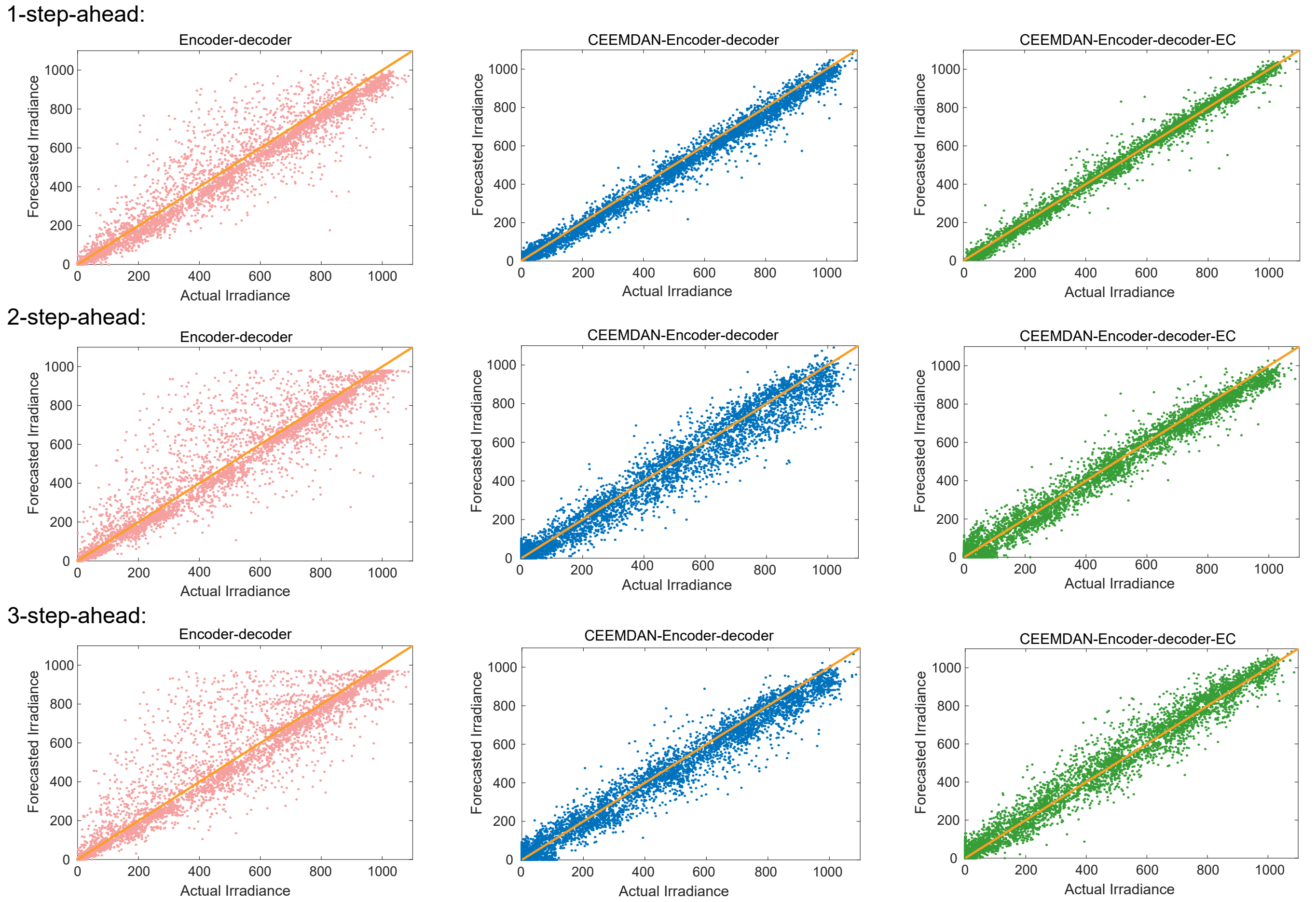

**Figure 4:** Scatter of the forecasting results of Hawaii dataset for 1-step ahead to 3-step ahead.

According to Fig. 4, there is a growing trend in distribution density of the scatter points, i.e., the similarity between the predicted values and actual values is increasing, which leads to the conclusion that the prediction capability of the model can be effectively enhanced by introducing serial decomposition and error compensation modules.

Table. 4 lists the evaluation metrics of each model, where RMSE, MAE, and nRMSE describe the degree of deviation between the predicted values and the actual values, and the correlation coefficient R describes the similarity between the predicted values and actual values. The best values for each group of comparison experiments have been highlighted in bold. For one-step ahead forecasting on Hawaii dataset, the RMSE of the Encoder-decoder model, CEEMDAN-Encoder-decoder model,and CEEMDAN-Encoder-decoder-EC model are 61.597 W/m$^2$, 39.738 W/m$^2$, and 32.912 W/m$^2$, respectively; the MAE of the three model are 25.879 W/m$^2$, 23.078 W/m$^2$, and 19.398 W/m$^2$, respectively; the nRMSE of the three model are 24.311%, 15.684%, and 12.989%, respectively; the values of R for the three models are 0.9944, 0.9726, 0.9959, respectively. The evaluation metrics of the CEEMDAN-Encoder-decoder model and CEEMDAN-Encoder-decoder-EC model for one-step-ahead prediction are better than the baseline model. After the introduction of the CEEMDAN module, the RMSE and MAE reduce by 35.4%, 10.8%, respectively. Compared with the Encoder-decoder model, the RMSE and MAE of the CEEMDAN-EC reduce by 46.6%, 25.0%, respectively. For the Oak Ridge dataset and the Alice Spring dataset presented in Table. 4, the evaluation metrics of the CEEMDAN-Encoder-decoder model and CEEMDAN-Encoder-decoder-EC model for one-step-ahead prediction are better than the baseline model. Introducing a dynamic error compensation module to the CEEMDAN-Encoder-decoder model, the performance of one-step-ahead prediction is improved, where in the Oak Ridge dataset, the RMSE is 34.433, the MAE is 22.9008, the nRMSE is 13.332%, and the correlation coefficient R is 0.9963. In the Alice Spring dataset, the RMSE reaches 34.433, the MAE reaches 22.9008, the nRMSE reaches 13.332%, and the correlation coefficient R reaches 0.9963.

For the two-step ahead and three-step ahead predictions, it can also be observed that the forecasting performance can be improved by introducing the decomposition module and dynamic error compensation module to the baseline model. Besides, similar phenomena can be observed on the Oak and Alice datasets. According to Table. 4, the CEEMDAN-Encoder-decoder model and the CEEMDAN-Encoder-decoder-EC model are better than the baseline model for one-step to three-step forecasting tasks to some extent.

In addition, Fig. 5 provides an overview of the changes in the metrics of each model through bar graphs, where the horizontal axis indicates the forecasting step and the vertical axis indicates the metrics. It can be seen that the metrics RMSE, MAE, and nRMSE generally present decreasing trend with the introduction of the decomposition module and the dynamic error compensation module at different time scales.

### 3.5. Comparison experiments

In this part, several comparison experiments are designed to validate the effectiveness of the proposed method. The experiments are divided into the comparison with different feature extraction modules and the comparison with hybrid models, which are used to evaluate the

**Table 4**
Evaluation metrics of ablation experiments.

| Model | Hawaii dataset | | | | Oak Ridge dataset | | | | Alice Spring dataset | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | RMSE | MAE | nRMSE(%) | R | RMSE | MAE | nRMSE(%) | R | RMSE | MAE | nRMSE(%) | R |
| 1-step ahead | | | | | | | | | | | | |
| Encoder-decoder | 61.597 | 25.879 | 24.311 | 0.9819 | 54.322 | 23.683 | 35.463 | 0.9726 | 83.630 | 31.182 | 32.381 | 0.9718 |
| C-Encoder-decoder | 39.738 | 23.078 | 15.684 | 0.9944 | 34.629 | **20.976** | 22.607 | **0.9950** | 38.214 | 24.878 | 14.843 | 0.9946 |
| C-Encoder-decoder-EC | **32.912** | **19.398** | **12.989** | **0.9959** | **30.693** | 21.840 | **20.037** | 0.9932 | **34.433** | **22.900** | **13.332** | **0.9963** |
| 2-step ahead | | | | | | | | | | | | |
| Encoder-decoder | 76.573 | 34.571 | 31.709 | 0.9720 | 70.084 | 31.252 | 45.753 | 0.9544 | 86.479 | 36.402 | 33.523 | 0.9697 |
| C-Encoder-decoder | 44.345 | 27.728 | 18.389 | 0.9915 | 36.848 | 20.311 | 24.055 | 0.9906 | 53.441 | 30.286 | 20.780 | 0.9893 |
| C-Encoder-decoder-EC | **38.503** | **22.385** | **15.966** | **0.9931** | **34.947** | **17.621** | **22.815** | **0.9889** | **47.893** | **25.037** | **18.623** | **0.9908** |
| 3-step ahead | | | | | | | | | | | | |
| Encoder-decoder | 83.999 | 35.053 | 34.838 | 0.9647 | 79.871 | 37.039 | 52.143 | 0.9405 | 98.320 | 41.318 | 38.458 | 0.9601 |
| C-Encoder-decoder | 50.119 | 31.068 | 20.783 | 0.9895 | 40.306 | 20.964 | 26.313 | 0.9860 | 66.542 | 38.203 | 25.876 | 0.9824 |
| C-Encoder-decoder-EC | **46.794** | **29.355** | **19.407** | **0.9902** | **39.240** | **22.220** | **25.617** | **0.9881** | **56.316** | **25.241** | **21.900** | **0.9872** |

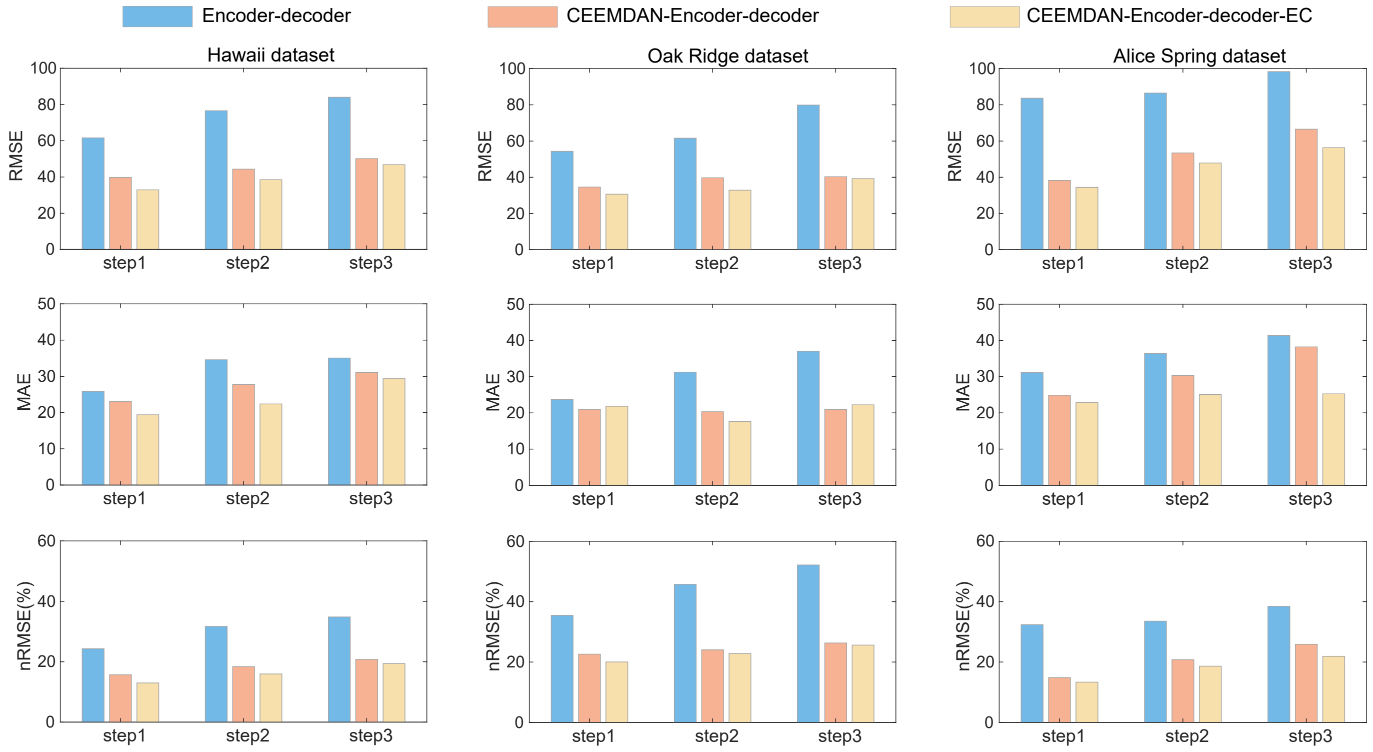

**Figure 5:** Bar graphs of evaluation metrics results on three datasets.

proposed encoder-decoder model and the hybrid deep learning model, respectively.

### 3.5.1. *Comparison with different feature extraction modules*

In order to assess the forecasting ability of the proposed encoder-decoder model, several comparison experiments based on deep learning methods are conducted in this part. The involved models consist of the MLP model, the LSTM model, the TCN model, and the proposed encoder-decoder model. The MLP is a back-propagation based fully connected neural network with a single hidden layer structure in the experiments. The LSTM is a two-layer structure whose output is dimensionally transformed through a fully connected layer. The TCN model consists of 6 residual blocks and a fully connected layer for dimension transformation. The hyperparameters can be obtained by grid search.

Figs. 6 to 8 present the one-step ahead forecasting performance of the above models on three datasets. Each figure contains two parts, namely the forecasted values curve and the scatter plot of forecasted values. From the irradiance curves of Figs. 6 to 8, it can be observed that all methods involved in the comparison experiments have great convergence performance and are able to forecast the variation of the actual irradiance to a certain extent. The distribution information of the actual values and the forecasted values can be obtained from the scatter graphs. The forecasted values and actual values of all the above models are distributed near the diagonal, where a more concentrated scatter distribution is observed for the proposed encoder-decoder model.

The detailed evaluation metrics including RMSE, nRMSE, MAE, and R of the above models for different forecasting horizons on three datasets are displayed in Table. 5. For one-step ahead forecasting in the Hawaii, Oak Ridge, and Alice Spring, the values of RMSE for the Encoder-decoder model are 61.597 W/m$^2$, 54.322 W/m$^2$, and 83.630 W/m$^2$, respectively; the values of MAE for the Encoder-decoder model are 25.879 W/m$^2$, 23.683 W/m$^2$, and 32.182 W/m$^2$, respectively; the values of nRMSE for the Encoder-decoder model are 24.311%, 35.643%, and 32.381%, respectively; the values of R for the Encoder-decoder model are 0.9819, 0.9726, 0.9718, respectively. Compared to the baseline models, the Encoder-decoder's RMSE decreases by 2.16%-6.65%,

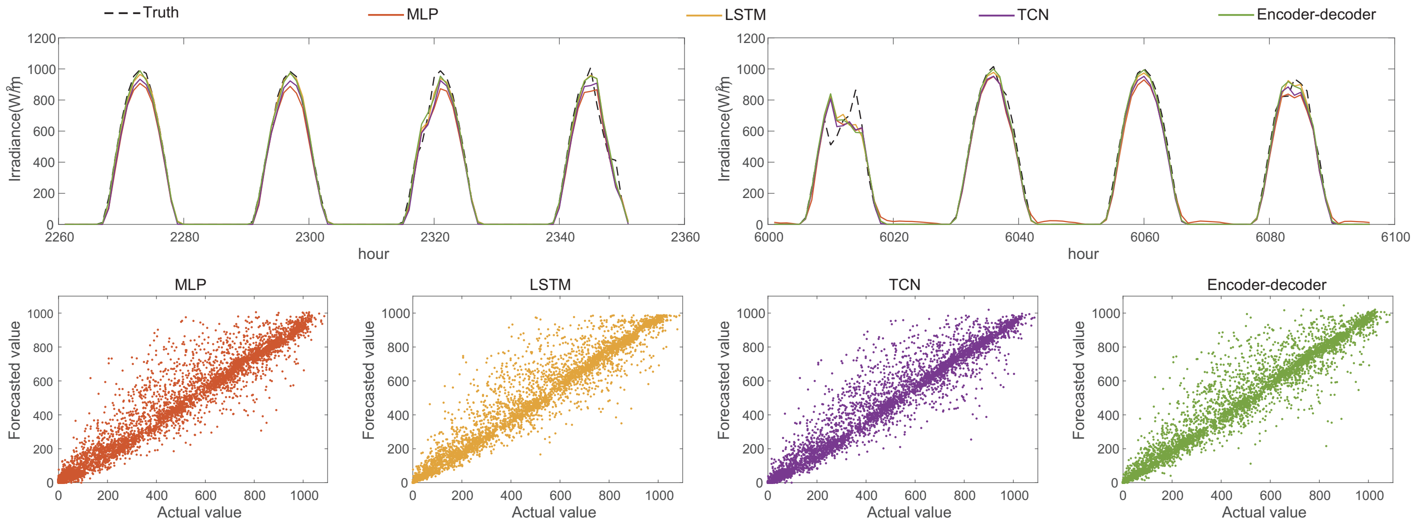

Figure 6: Results of one-step ahead forecasting on Hawaii dataset.

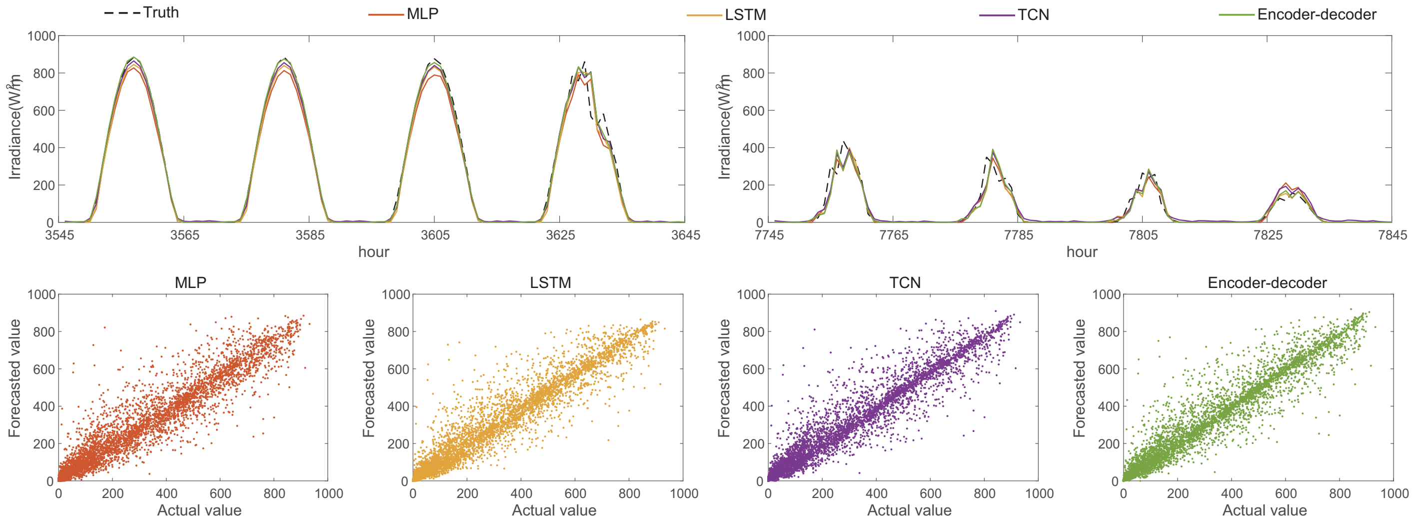

Figure 7: Results of one-step ahead forecasting on Oak Ridge dataset.

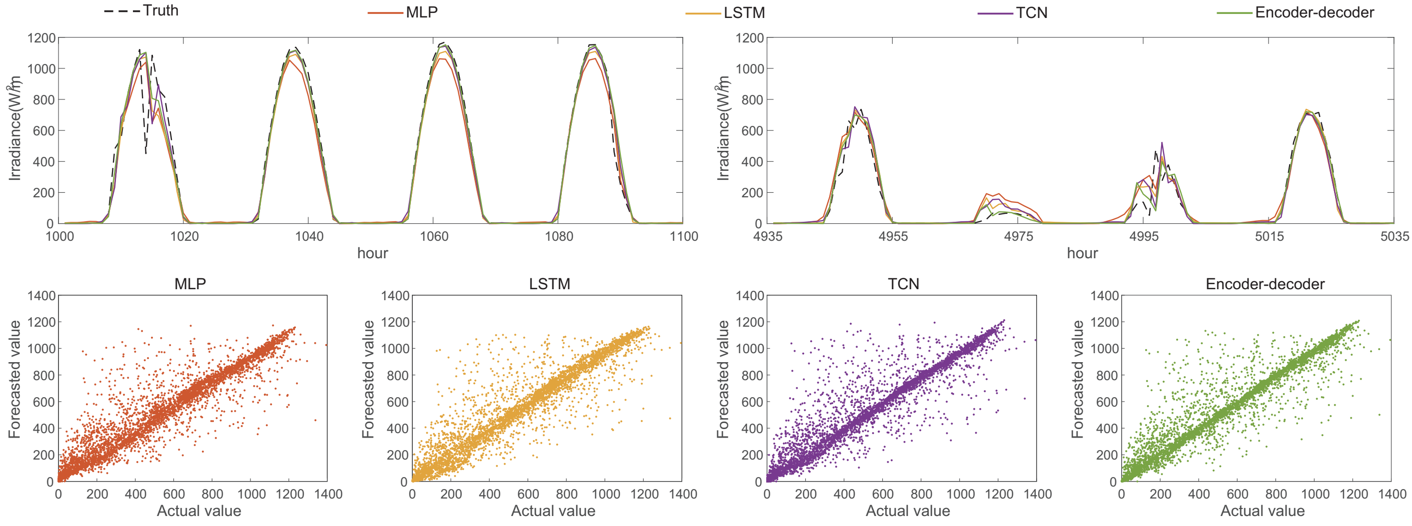

Figure 8: Results of one-step ahead forecasting on Alice Spring dataset.

**Table 5**
Forecasting performance of different feature extraction modules.

| Model | 1-step ahead |  |  |  | 2-step ahead |  |  |  | 3-step ahead |  |  |  |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
|  | RMSE | MAE | nRMSE(%) | R | RMSE | MAE | nRMSE(%) | R | RMSE | MAE | nRMSE(%) | R |
| Hawaii dataset |  |  |  |  |  |  |  |  |  |  |  |  |
| MLP | 65.982 | 35.518 | 26.042 | 0.9798 | 88.504 | 53.191 | 36.699 | 0.9649 | 104.01 | 64.577 | 43.127 | 0.9594 |
| LSTM | 63.741 | 29.604 | 26.431 | 0.9810 | 78.003 | 36.926 | 32.345 | 0.9708 | 85.484 | 43.580 | 35.447 | 0.9650 |
| TCN | 62.956 | 30.687 | 24.848 | 0.9814 | 77.018 | 38.104 | 31.936 | 0.9716 | 84.449 | 42.965 | 35.018 | **0.9659** |
| Encoder-decoder | **61.597** | **25.879** | **24.311** | **0.9819** | **76.573** | **34.571** | **31.709** | **0.9720** | **83.999** | **35.053** | **34.838** | 0.9647 |
| Oak Ridge dataset |  |  |  |  |  |  |  |  |  |  |  |  |
| MLP | 59.448 | 28.830 | 38.820 | 0.9672 | 79.178 | 43.923 | 51.691 | 0.9413 | 91.457 | 51.334 | 59.707 | 0.9230 |
| LSTM | 55.470 | 25.787 | 36.223 | 0.9711 | 72.026 | 34.253 | 47.021 | 0.9510 | 82.795 | 39.360 | 54.052 | 0.9353 |
| TCN | 55.168 | 28.033 | 36.025 | 0.9716 | 71.245 | 34.123 | 46.511 | 0.9519 | 81.518 | 39.216 | 53.218 | 0.9368 |
| Encoder-decoder | **54.322** | **23.683** | **35.463** | **0.9726** | **70.084** | **31.252** | **45.753** | **0.9544** | **79.871** | **37.039** | **52.143** | **0.9405** |
| Alice Spring dataset |  |  |  |  |  |  |  |  |  |  |  |  |
| MLP | 89.471 | 43.682 | 34.643 | 0.9683 | 98.612 | 50.934 | 38.302 | 0.9628 | 105.08 | 54.885 | 40.817 | 0.9561 |
| LSTM | 85.713 | 35.538 | 33.188 | 0.9703 | 92.602 | 40.092 | 35.967 | 0.9651 | 100.08 | 45.514 | 38.878 | 0.9593 |
| TCN | 88.472 | 36.012 | 34.256 | 0.9684 | 95.218 | 42.147 | 36.984 | 0.9635 | 100.71 | 45.474 | 39.122 | 0.9586 |
| Encoder-decoder | **83.630** | **31.182** | **32.381** | **0.9718** | **86.479** | **36.402** | **33.523** | **0.9697** | **98.320** | **41.318** | **38.458** | **0.9601** |

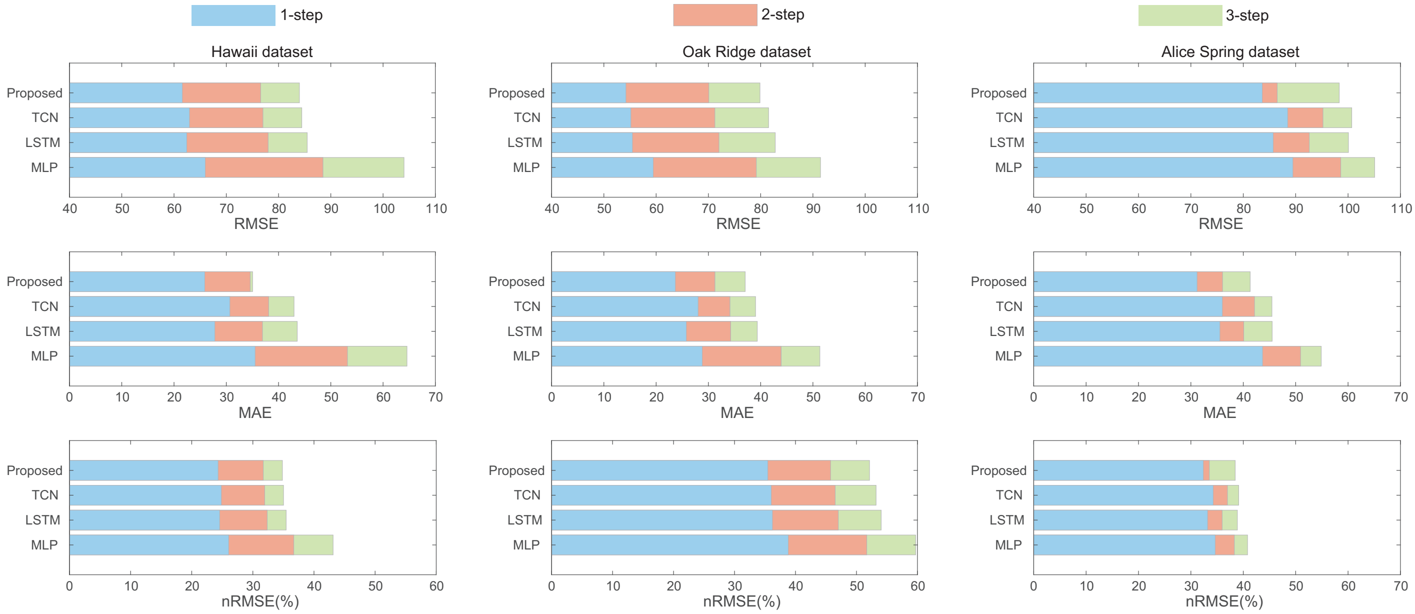

**Figure 9:** Stacked bar graphs for evaluation indicators of 1-step ahead to 3-step ahead.

1.53%-8.62%, and 2.43%-6.53% for the three datasets, respectively. The Encoder-decoder decreases MAE by 2.43%-27.14%, 8.16%-17.85%, and 12.26%-28.62% compared to the baseline models on the three datasets, respectively. For the two-step ahead forecasting, the Encoder-decoder model reduces the RMSE and MAE by 0.58%-13.48%, 6.38%-34.77% on the Hawaii dataset, respectively; the Encoder-decoder model reduces the RMSE and MAE by 1.63%-11.40%, 8.41%-28.85% on the Oak Ridge dataset, respectively; the Encoder-decoder model reduces the RMSE and MAE by 6.61%-12.30%, 9.20%-28.63% on the Alice Spring dataset, respectively. For the three-step ahead forecasting, the forecasting performance of the Encoder decoder model is also improved compared to the baseline models on the three datasets.

In addition, the stacked bar graphs in Fig. 9 provide a more visual manner to illustrate the forecasting performance of the models over different forecasting horizons. The horizontal axis in the figure represents the numerical value of the evaluation indicators, the vertical axis marks the models participating in the comparison experiments, and the different colors of the bar graph represent different forecasting horizons. According to the table and the stacked bar graphs, the performance of the MLP model generates the maximum forecasting error in the comparison experiments of each step size. The LSTM model and the TCN model perform close to each other under different datasets and different stride lengths. It is noteworthy that the proposed encoder-decoder model has a certain performance improvement over the other models, which achieves the minimum forecasting errors in each comparison experiment.

The above comparison experiments and corresponding analysis indicate that the proposed model has great capability to extract temporal features. The results support the feasibility of introducing other modules on the basis of the encoder-decoder to a certain extent.

### *3.5.2. Performance evaluation for different hybrid models*

In this section, the proposed hybrid deep learning model based on encoder-decoder is compared with other different forecasting techniques. For comparison experiments of the hybrid model, signal decomposition module is applied to each model (i.e.,the MLP, the LSTM,

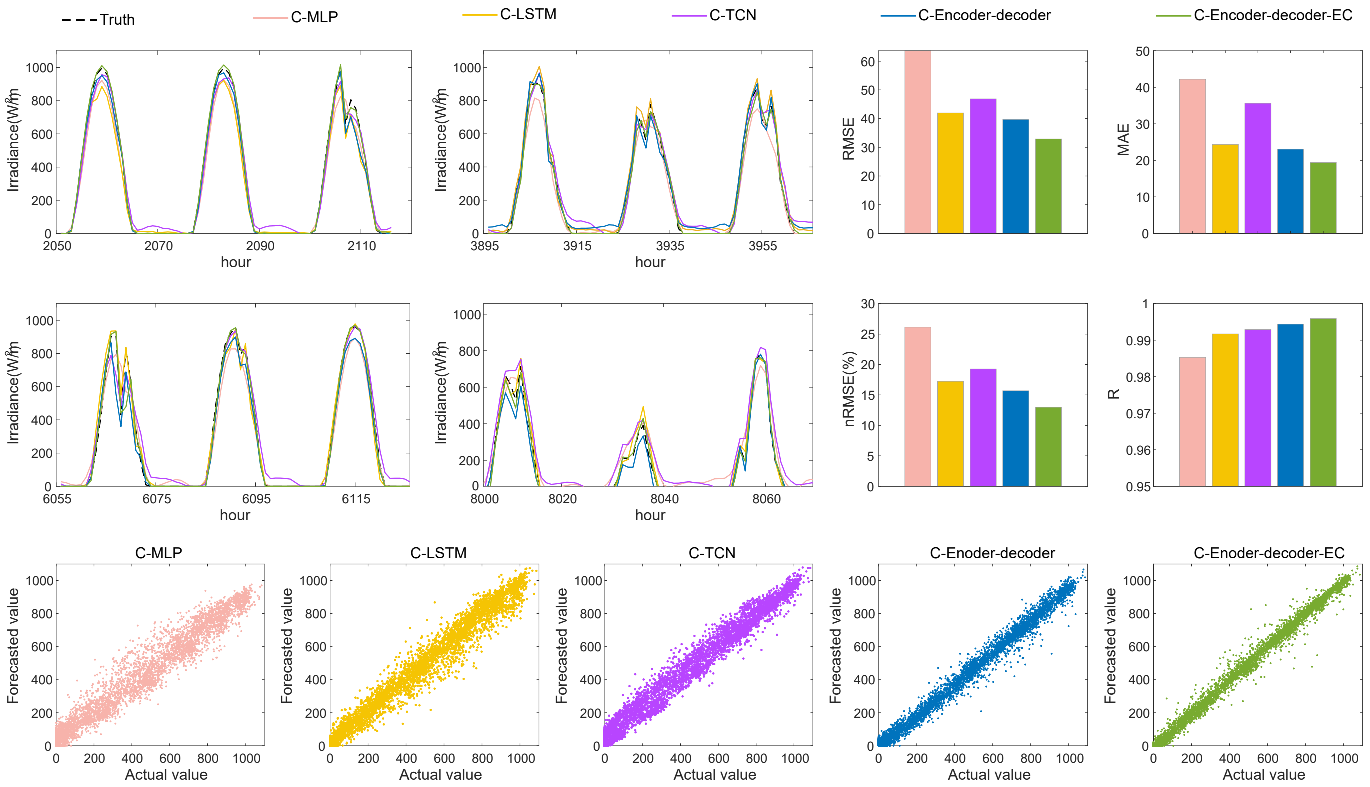

**Figure 10:** Forecasting result and evaluation metrics of 1-step ahead on Hawaii dataset.

and the TCN), where the baseline models are comprised of CEEMDAN-based MLP (C-MLP), CEEMDAN-based LSTM (C-LSTM), CEEMDAN-based TCN (C-TCN), CEEMDAN-based Encoder-decoder (C-Encoder-decoder). The structure of each hybrid model is the same as the baseline model in the previous section.

The one-step ahead forecasted results on three datasets are plotted in Figs. 10 to 12, which consist of the forecast curves, the bar graphs of the evaluation indicators, and the scatter graphs of forecasted values. The forecast curves show the forecasted value of irradiance for four seasons, where 70 hours are selected for each season. According to the scatter plot, the deviation between the actual value and the forecasted value can be intuitively observed. The bar graphs present the variation of the evaluation metrics of the above models. To comprehensively evaluate the proposed hybrid model, comparison experiments with multi-step prediction are executed. The results are shown in Fig. 13. Moreover, the assessment indicators of the models are depicted in Table. 6.

From Figs. 10 to 12 and Table. 6, it can be observed that:

(1) The proposed model shows the best forecasting performance compared to other hybrid models. For one-step ahead forecasting in the Hawaii, Oak Ridge, and Alice Spring, the RMSE of the CEEMDAN-Encoder-decoder-EC model are 32.9127 W/m$^2$, 30.693 W/m$^2$, and 34.433 W/m$^2$, respectively; the MAE of Encoder-decoder model are 19.398 W/m$^2$, 21.840 W/m$^2$, and 22.900 W/m$^2$, respectively; the nRMSE of the CEEMDAN-Encoder-decoder-EC model are 12.989%, 20.037%, and 13.332%, respectively; the R of the CEEMDAN-Encoder-decoder-EC model are 0.9959, 0.9932, 0.9963, respectively. For two-step ahead forecasting in the Hawaii, Oak Ridge, and Alice Spring, the RMSE of the CEEMDAN-Encoder-decoder-EC model are 38.503 W/m$^2$, 34.947 W/m$^2$, and 47.893 W/m$^2$, respectively; the MAE of Encoder-decoder model are 22.385 W/m$^2$, 17.621 W/m$^2$, and 25.037 W/m$^2$, respectively; the nRMSE of the CEEMDAN-Encoder-decoder-EC model are 15.966%, 22.815%, and 18.623%, respectively; the R of the CEEMDAN-Encoder-decoder-EC model are 0.9931, 0.9889, 0.9908, respectively. Besides,the performance metrics of the C-Encoder-decoder-EC model are considerably better than the others in the three-step forecasting tasks including two-step ahead and three-step ahead. For instance, the RMSE of the proposed C-Encoder-decoder-EC method is reduced by 49.26%, 25.12%, 10.76%, and 9.96% in the three-step ahead forecasting task for the Alice Spring dataset relative to C-MLP model, C-LSTM model, C-TCN model, and C-Encoder-decoder model, respectively.

(2) All the hybrid models involved in the comparison experiments performed well on the one-step ahead forecasting task, but accuracy decreased as the number of prediction steps increased. However, the proposed model still performs best for the multi-step forecasting task. Overall, the predicted values of the proposed C-Encoder-decoder-EC model have the lowest degree of deviation across all datasets, indicating that the predicted values of the model are highly consistent with the actual values. According to indicators bar graphs and evaluation metrics table, the C-Encoder-decoder-EC model shows better forecasting performance than the other hybrid models. The excellent performance on multi-step forecasting tasks on different datasets indicates that the proposed model has great robustness.

(3) From a seasonal perspective, the proposed method can effectively adapt to seasonal changes. According to the forecasting curve, the forecasting performance of the model does not decrease significantly with the seasonal changes. Furthermore, the forecasting curves indicate that the proposed model is effective at forecasting

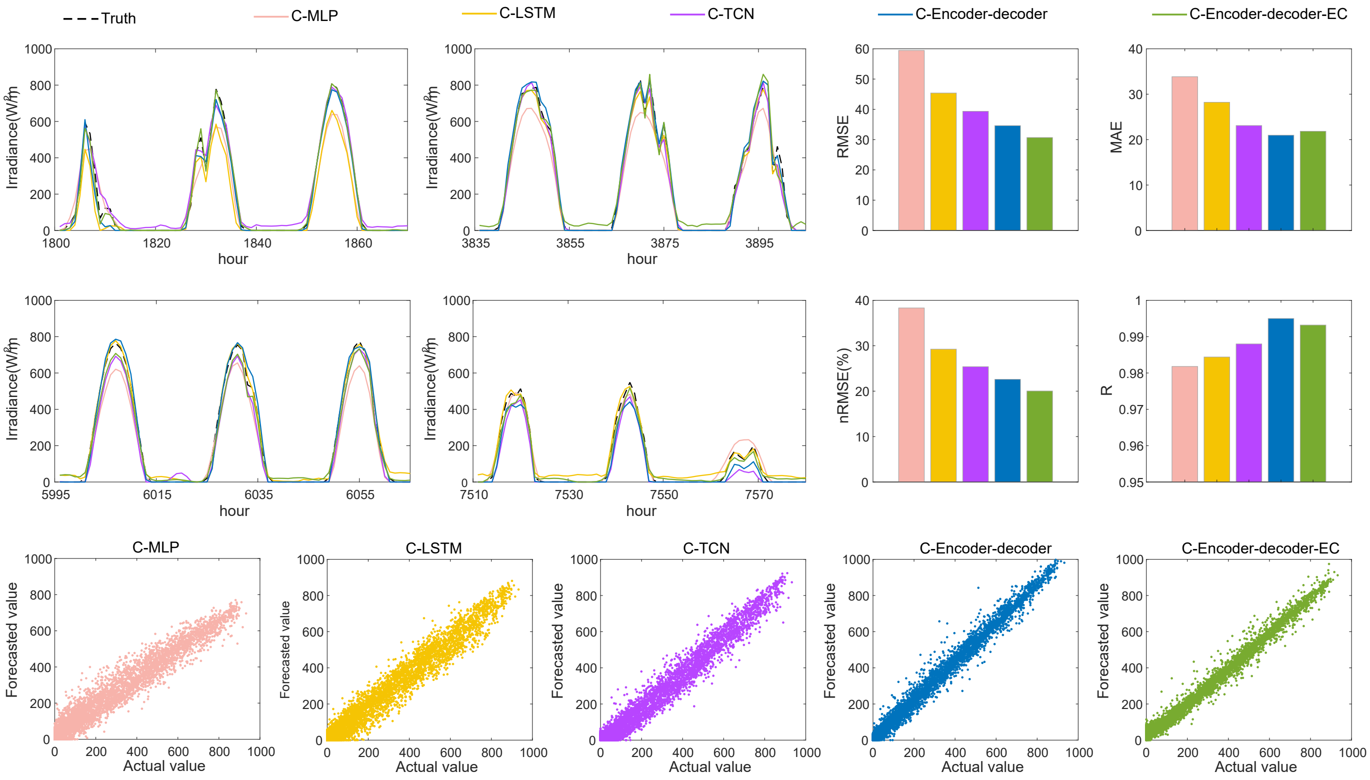

**Figure 11:** Forecasting result and evaluation metrics of 1-step ahead on Oak Ridge dataset.

irradiance fluctuations. Compared to other methods, the proposed model shows the best irradiance tracking performance. It demonstrates that the proposed model based on encoder-decoder has better ability to extract inherent temporal features.

Based on the above observations, it can be concluded that the proposed encoder-decoder structure can extract the hidden representation of the temporal features effectively, meanwhile, the decomposition module and the dynamic error compensation method can effectively enhance the predictive performance of the model.

## 3.6. Further exploration

In this section, we explore the proposed model more extensively. The computational cost of the model and the effect of noise on the model are discussed separately.

### *3.6.1. Computational cost*

To explore the practicality of the proposed method, the computational cost of the model is discussed in this section. We describe the computational cost of the model from two perspectives of space consumption and time consumption. The space consumption is expressed by the number of parameters. The time consumption is described by floating-point operations (FLOPs)[51] and runtime. The metric of FLOPs represents Floating-Point Operations, which is often used to measure the computational power consumption of the model. The FLOPs can be used as an indirect metric to measure the speed of neural network models. The running time is divided into training time for the whole dataset and testing time for a single moment.

We discuss the computational cost of the models using the Hawaii dataset as an example. All models are trained with a setting of 150 epochs and a batch size of 16. The details of the computational cost can be found in Table. 7. The testing time represents the predicted time consumed by each batch.

According to the Table. 7, the number of parameters for C-LSTM, C-TCN, C-Encoder-decoder, and C-Encoder-decoder-EC are 291K, 591K, 954K, and 954K respectively. The C-LSTM model has the lowest FLOPs, meaning it requires the least amount of computation. The CEEMDAN-Encoder-decoder-EC model has the longest training time of 16.5 hours. Besides, the testing time for all models are well below the forecasting interval.

Although the proposed C-Encoder-decoder-EC model has the largest time consumption during the testing stage, the time consumption is far less than the forecasting interval for the hourly forecasting task, which does not affect the practical applications. For an actual forecasting task, we pay more attention to the forecasting accuracy of the model. The proposed method provides a higher forecasting accuracy, which is more instructive for tasks such as pv grid integration and real-time dispatch.

### *3.6.2. Noise effect*

In practical applications, noises often derive from data collection processes and covariates.

The effects of noises from data collection such as sensor jitter can be effectively mitigated by data preprocessing. In practice, irradiance data and the corresponding covariates are mostly collected at minute resolution, such as the three datasets in this study which are sampled at 5 minute intervals. For hourly irradiance prediction, averaging the data collected in the hourly range is a reliable way to reduce the effect of noise.

In addition, the prediction model brings noises when introducing covariates. Since the uncorrelated components of the covariate series are not helpful for prediction, it can be considered that the lower the covariate correlation with irradiance, the greater the proportion of covariate noise. Accordingly, the amount of useful information for irradiance prediction in the covariate series can be measured by the correlation coefficient. An extreme example is that the correlation coefficient between

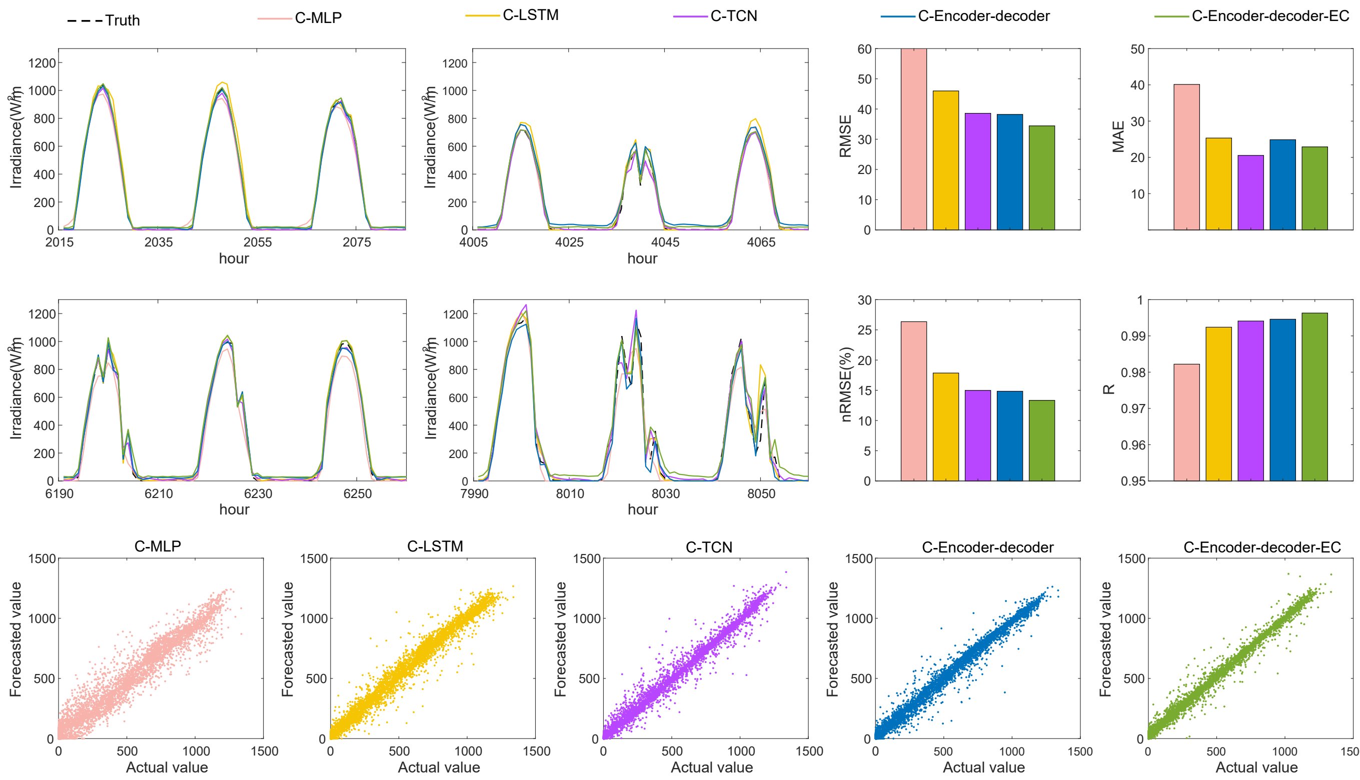

**Figure 12:** Forecasting result and evaluation metrics of 1-step ahead on Alice Spring dataset.

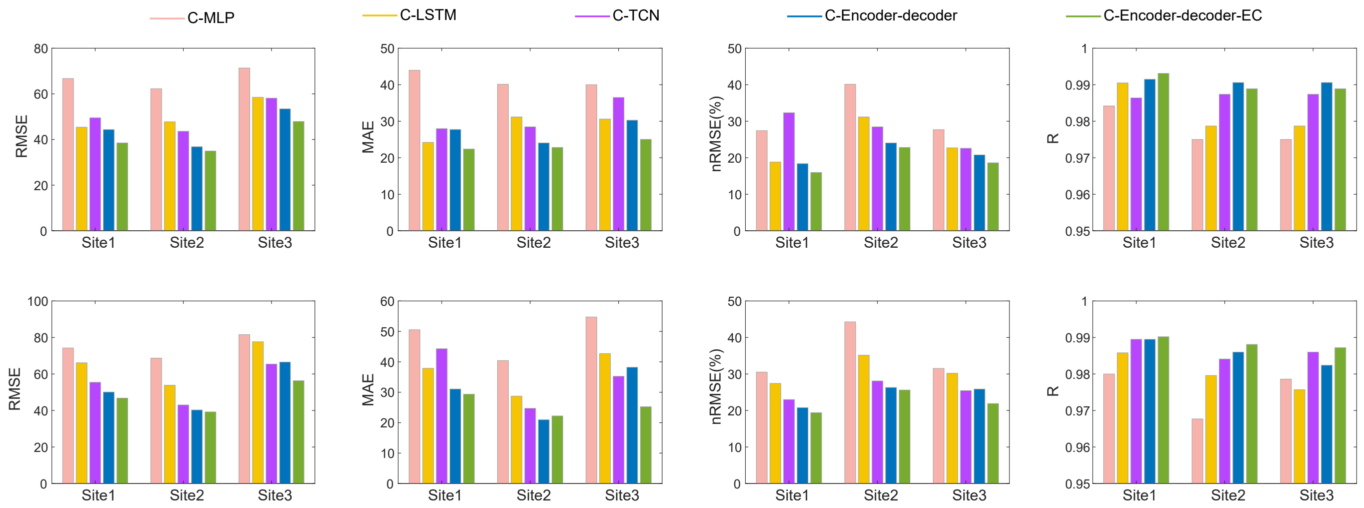

**Figure 13:** Results of multi-step ahead forecasting tasks, where first row represents 2-step ahead forecasting task and second row represents 3-step ahead forecasting task.

covariates and irradiance is close to zero when the covariate is white noise. Consequently, when the covariates are improperly chosen, there will be unnecessary noise introduced. For example, the precipitation is one of the factors affecting irradiance, but for prediction tasks at the hour scale, the precipitation changes discontinuously, making it difficult to provide regular information.

Furthermore, the DEC mechanism is based on the assumption that the predictable components are not sufficiently extracted by the prediction model, which implies there is valuable information in the error information. It's worth noting that the error compensation mechanism is proposed for mining existing information, which does not increase the amount of information. Therefore, if the proportion of noise in the error information is too high, the predictable components will be diluted, resulting in difficulty for the model to extract the corresponding features. Since a well-trained deep learning model is usually able to establish a stable mapping between irradiance and covariates, the noise in the error information may originate from the unpredictable components in the irradiance sequence. Reasons for the presence of such unpredictable components include, for example, the randomness of the sequence itself, insufficient information on the covariates, and limitations of the

**Table 6**
Forecasting performance of different hybrid models.

| Dataset | Forecasting step | Evaluation metrics | Model | | | | |
|---|---|---|---|---|---|---|---|
| | | | C-MLP | C-LSTM | C-TCN | C-Encoder-decoder | C-Encoder-decoder-EC |
| Hawaii dataset | 1-step ahead | RMSE | 63.634 | 41.972 | 46.875 | 39.738 | **32.912** |
| | | MAE | 42.227 | 24.370 | 35.646 | 23.078 | **19.398** |
| | | nRMSE(%) | 26.151 | 17.249 | 19.264 | 15.684 | **12.989** |
| | | R | 0.9853 | 0.9917 | 0.9929 | 0.9944 | **0.9959** |
| | 2-step ahead | RMSE | 66.716 | 45.406 | 49.531 | 44.345 | **38.503** |
| | | MAE | 43.959 | 24.216 | 27.975 | 27.728 | **22.385** |
| | | nRMSE(%) | 27.418 | 18.828 | 32.336 | 18.389 | **15.966** |
| | | R | 0.9842 | 0.9905 | 0.9864 | 0.9915 | **0.9931** |
| | 3-step ahead | RMSE | 74.267 | 66.148 | 55.450 | 50.119 | **46.794** |
| | | MAE | 50.540 | 37.880 | 44.298 | 31.068 | **29.355** |
| | | nRMSE(%) | 30.521 | 27.430 | 22.994 | 20.783 | **19.407** |
| | | R | 0.9800 | 0.9858 | 0.9895 | 0.9895 | **0.9902** |
| Oak Ridge dataset | 1-step ahead | RMSE | 59.456 | 45.387 | 39.352 | 34.629 | **30.693** |
| | | MAE | 33.853 | 28.224 | 23.104 | **20.976** | 21.840 |
| | | nRMSE(%) | 38.319 | 29.252 | 25.396 | 22.607 | **20.037** |
| | | R | 0.9818 | 0.9844 | 0.9880 | **0.9950** | 0.9932 |
| | 2-step ahead | RMSE | 62.260 | 47.768 | 43.609 | 36.848 | **34.947** |
| | | MAE | 35.350 | 27.070 | 24.525 | 20.311 | **17.621** |
| | | nRMSE(%) | 40.126 | 31.185 | 28.470 | 24.055 | **22.815** |
| | | R | 0.9750 | 0.9787 | 0.9874 | **0.9906** | 0.9889 |
| | 3-step ahead | RMSE | 68.694 | 53.850 | 43.047 | 40.306 | **39.240** |
| | | MAE | 40.419 | 28.725 | 24.706 | **20.964** | 22.220 |
| | | nRMSE(%) | 44.273 | 35.156 | 28.102 | 26.313 | **25.617** |
| | | R | 0.9677 | 0.9796 | 0.9841 | 0.9860 | **0.9881** |
| Alice Spring dataset | 1-step ahead | RMSE | 67.868 | 45.984 | 38.585 | 38.214 | **34.433** |
| | | MAE | 40.115 | 25.331 | 20.539 | 24.878 | **22.900** |
| | | nRMSE(%) | 26.352 | 17.855 | 14.982 | 14.843 | **13.332** |
| | | R | 0.9822 | 0.9924 | 0.9941 | 0.9946 | **0.9963** |
| | 2-step ahead | RMSE | 71.320 | 58.504 | 58.127 | 53.441 | **47.893** |
| | | MAE | 40.005 | 30.603 | 36.504 | 30.286 | **25.037** |
| | | nRMSE(%) | 27.692 | 22.724 | 22.602 | 20.780 | **18.623** |
| | | R | 0.9795 | 0.9863 | 0.9880 | 0.9893 | **0.9908** |
| | 3-step ahead | RMSE | 81.578 | 77.672 | 65.443 | 66.542 | **56.316** |
| | | MAE | 54.728 | 42.719 | 35.242 | 38.203 | **25.241** |
| | | nRMSE(%) | 31.525 | 30.205 | 25.449 | 25.876 | **21.900** |
| | | R | 0.9786 | 0.9757 | 0.9860 | 0.9824 | **0.9872** |

**Table 7**
Computational cost of the proposed method and baseline methods. An example of the experiment with one-step ahead prediction of the Hawaii dataset (batch size = 16, epochs = 150).

| Model | Parameters | FLOPs | Testing | Training |
|---|---|---|---|---|
| C-LSTM | 291K | 82.38M | 16.8ms | 1.74h |
| C-TCN | 591K | 1.635G | 19.9ms | 2.52h |
| C-Encoder-decoder | 954K | 1.740G | 28.1ms | 3.13h |
| C-Encoder-decoder-EC | 954K | 1.740G | 146ms | 16.5h |

model's feature extraction capabilities. Therefore, insufficient covariate information will result in a large proportion of noise in the input data, which will affect the performance of the DEC. To specify this effect, we analyzed the ablation experiment for abnormal results.

As discussed in section 3.3 in main body, the nRMSE is able to compare model performance across datasets in a scale-free manner. Similarly, we normalize the MAE to get a scale-free metric nMAE to facilitate the comparison of model performance across datasets. We then compare the performance of each benchmark model under different datasets. Table. 4 shows the nRMSE of each model on the three datasets with one-step ahead prediction, two-step ahead prediction, and three-step ahead forecasting, respectively. The nMAE of the models can be seen in Table. 8.

The results show that all models have the worst prediction performance on the Oak Ridge dataset, compared with the other datasets. The phenomenon indicates that there is an inherent difficulty in predicting the Oak Ridge dataset. The reason for the phenomenon may be that the selected covariates are not the most appropriate for the Oak Ridge dataset, which brings additional noises.

Only the experiment on Oak dataset shows a slight increase in MAE/nMAE with the introduction of the DEC. This indicates that the noise impact of the dataset makes the proportion of noise components in the error information increase, resulting in the dilution of predictable components in the error information, making feature extraction difficult. Besides, the RMSE/nRMSE is reduced after the introduction of the dynamic error compensation mechanism, which indicates that DEC is

**Table 8**
The metric of nMAE at Hawaii / Alice Spring / Oak Ridge (Bold indicates **worst** performance).

| Model | one-step ahead(%) | two-step ahead(%) | three-step ahead(%) |
|---|---|---|---|
| C-MLP | 17.60/15.81/**21.32** | 18.32/15.77/**22.26** | 21.07/21.57/**25.45** |
| C-LSTM | 10.16/9.98/**17.77** | 10.09/2.06/**17.04** | 15.79/16.84/**18.09** |
| C-TCN | 14.64/9.10/**16.38** | 11.66/14.39/**15.44** | **18.36**/13.89/16.13 |
| C-Encoder-decoder | 9.62/9.81/**13.21** | 11.56/11.94/**12.79** | 12.95/14.61/**13.68** |
| C-Encoder-decoder-EC | 8.09/9.03/**13.75** | 9.33/9.87/**11.09** | 12.24/9.95/**13.99** |

still able to reasonably compensate for outliers under the influence of dataset noise.

To sum up, we discuss two different types of noise, introduced by the data collection process and introduced by covariates, respectively. The noise introduced by data collection is easily mitigated by intra-hour averaging and has little impact on the model. While the noise introduced by covariates is mainly due to inappropriate covariate selection and can be mitigated by introducing more correlated covariates, such as the numerical weather prediction (NWP) information.

# 4. Conclusion

Solar irradiance is a key factor impacting PV power generation, therefore accurate irradiance forecasting is vital to the utilization of solar resources. In this paper, a hybrid deep learning model was proposed for hourly solar irradiance forecasting, where the model is comprised of a series decomposition modules, encoder-decoder module, and error compensation module. The most noteworthy advantages of the proposed method can be summed up as follows: (a) The introduction of the series decomposition modules can effectively reduce the non-stability of irradiance sequences, as well as effectively separate out the component that is hard to predict. (b) The novel encoder-decoder module has excellent temporal feature extraction capability due to the cascade structure. (c) An end-to-end model is utilized for error compensation, enabling prediction performance improvements without increasing model parameters. (d) A new loss function is designed for multi-objective optimization to balance output performance over multiple time steps.

To validate the model performance, multiple sets of comparison experiments on three public datasets were designed. Hourly forecasting experiment on Hawaii dataset shows that the RMSE, MAE, nRMSE, and R of the proposed C-Encoder-decoder-EC model are 32.912 W/m$^2$, 19.398 W/m$^2$, 12.989%, and 0.9959, respectively. Compared with benchmark models, the RMSE and MAE of the proposed model are reduced by more than 21.59% and 20.40%, respectively. For Alice Spring dataset, the RMSE and MAE of the proposed model are reduced by more than 10.76% and 20.40%, respectively. For Oak Ridge dataset, the hourly forecasting experiment indicates that RMSE and MAE of the proposed model are reduced by more than 22.00% and 5.47%, respectively. The experimental results validate the effectiveness of the proposed model and demonstrate better prediction performance of the proposed model than the benchmark models without considering the effect of computational cost. Besides, the positive results suggest that the proposed method is a reliable alternative to hourly irradiance forecasting.

Moreover, only historical meteorological information is considered as exogenous input in this work, where the future meteorological information is unknown. Ignoring the covariates corresponding to the future moment causes the prediction performance of the models to be limited by the lack of information. NWP is a method to provide information about future covariates, which deserves to be accorded great attention. In future work, we will consider the combination of deep learning models with the NWP information to enhance the input features.

# Acknowledgements

This work was supported in part by the Zhishan Young Scholar Program of Southeast University; in part by the Fundamental Research Funds for the Central Universities under Grant 2242021R41118. Besides, we thank the Big Data Computing Center of Southeast University for providing the facility support on the numerical calculations.

# References

[1] E. Kabir, P. Kumar, S. Kumar, A. A. Adelodun, and K.-H. Kim, “Solar energy: Potential and future prospects,” *Renewable and Sustainable Energy Reviews*, vol. 82, pp. 894–900, 2018. https://doi.org/10.1016/j.rser.2017.09.094.

[2] A. S. B. M. Shah, H. Yokoyama, and N. Kakimoto, “High-precision forecasting model of solar irradiance based on grid point value data analysis for an efficient photovoltaic system,” *IEEE Transactions on Sustainable Energy*, vol. 6, pp. 474–481, 2015. https://doi.org/10.1109/TSTE.2014.2383398.

[3] REN21, Renewables 2020 global status report. https://www.ren21.net/wp-content/uploads/2019/05/GSR2021_Full_Report.pdf.

[4] M. Guermoui, F. Melgani, K. Gairaa, and M. L. Mekhalfi, “A comprehensive review of hybrid models for solar radiation forecasting,” *Journal of Cleaner Production*, vol. 258, p. 120357, 2020. https://doi.org/10.1016/j.jclepro.2020.120357.

[5] S. Sun, S. Wang, G. Zhang, and J. Zheng, “A decomposition-clustering-ensemble learning approach for solar radiation forecasting,” *Solar Energy*, vol. 163, pp. 189–199, 2018. https://doi.org/10.1016/j.solener.2018.02.006.

[6] D. S. Kumar, G. M. Yagli, M. Kashyap, and D. Srinivasan, “Solar irradiance resource and forecasting: a comprehensive review,” *IET Renewable Power Generation*, vol. 14, pp. 1641–1656, 2020. https://doi.org/10.1049/iet-rpg.2019.1227.

[7] C. Voyant, G. Notton, S. Kalogirou, M.-L. Nivet, C. Paoli, F. Motte, and A. Fouilloy, “Machine learning methods for solar radiation forecasting: A review,” *Renewable Energy*, vol. 105, pp. 569–582, 2017. https://doi.org/10.1016/j.renene.2016.12.095.

[8] D. Yang, J. Kleissl, C. A. Gueymard, H. T. Pedro, and C. F. Coimbra, “History and trends in solar irradiance and pv power forecasting: A preliminary assessment and review using text mining,” *Solar Energy*, vol. 168, pp. 60–101, 2018. https://doi.org/10.1016/j.solener.2017.11.023.

[9] A. Murata, H. Ohtake, and T. Oozeki, “Modeling of uncertainty of solar irradiance forecasts on numerical weather predictions with the estimation of multiple confidence intervals,” *Renewable Energy*, vol. 117, pp. 193–201, 2018. https://doi.org/10.1016/j.renene.2017.10.043.

[10] M. Diagne, M. David, P. Lauret, J. Boland, and N. Schmutz, “Review of solar irradiance forecasting methods and a proposition for small-scale insular grids,” *Renewable and Sustainable Energy Reviews*, vol. 27, pp. 65–76, 2013. https://doi.org/10.1016/j.rser.2013.06.042.

[11] M. David, F. Ramahatana, P.-J. Trombe, and P. Lauret, “Probabilistic forecasting of the solar irradiance with recursive arma and garch models,” *Solar Energy*, vol. 133, pp. 55–72, 2016. https://doi.org/10.1016/j.solener.2016.03.064.

[12] H. Long, Z. Zhang, and Y. Su, “Analysis of daily solar power prediction with data-driven approaches,” *Applied Energy*, vol. 126, pp. 29–37, 2014. https://doi.org/10.1016/j.apenergy.2014.03.084.

[13] D. Yang, P. Jirutitijaroen, and W. M. Walsh, “Hourly solar irradiance time series forecasting using cloud cover index,” *Solar Energy*, vol. 86, pp. 3531–3543, 2012. https://doi.org/10.1016/j.solener.2012.07.029.

[14] L. Xie, J. Zhao, H. Wei, K. Zhang, and G. Pang, “Online kernel-based structured output svm for early expression detection,” *IEEE Signal Processing Letters*, vol. 26, no. 9, pp. 1305–1309, 2019. https://doi.org/10.1109/LSP.2019.2929713.

[15] L. Xie, D. Tao, and H. Wei, “Early expression detection via online multi-instance learning with nonlinear extension,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 5, pp. 1486–1496, 2018. https://doi.org/10.1109/TNNLS.2018.2869891.

[16] R. C. Deo, X. Wen, and F. Qi, “A wavelet-coupled support vector machine model for forecasting global incident solar radiation using limited

meteorological dataset," *Applied Energy*, vol. 168, pp. 568–593, 2016. https://doi.org/10.1016/j.apenergy.2016.01.130.

[17] I. A. Ibrahim and T. Khatib, "A novel hybrid model for hourly global solar radiation prediction using random forests technique and firefly algorithm," *Energy Conversion and Management*, vol. 138, pp. 413–425, 2017. https://doi.org/10.1016/j.enconman.2017.02.006.

[18] A. N. Celik and T. Muneer, "Neural network based method for conversion of solar radiation data," *Energy Conversion and Management*, vol. 67, pp. 117–124, 2013. https://doi.org/10.1016/j.enconman.2012.11.010.

[19] L. Xie, W. Guo, H. Wei, Y. Tang, and D. Tao, "Efficient unsupervised dimension reduction for streaming multiview data," *IEEE Transactions on Cybernetics*, vol. 52, no. 3, pp. 1772–1784, 2022. https://doi.org/10.1109/TCYB.2020.2996684.

[20] Y. Zhou, Y. Liu, D. Wang, X. Liu, and Y. Wang, "A review on global solar radiation prediction with machine learning models in a comprehensive perspective," *Energy Conversion and Management*, vol. 235, p. 113960, 2021. https://doi.org/10.1016/j.enconman.2021.113960.

[21] X. Huang, Q. Li, Y. Tai, Z. Chen, J. Zhang, J. Shi, B. Gao, and W. Liu, "Hybrid deep neural model for hourly solar irradiance forecasting," *Renewable Energy*, vol. 171, pp. 1041–1060, 2021. https://doi.org/10.1016/j.renene.2021.02.161.

[22] R. Prasad, M. Ali, Y. Xiang, and H. Khan, "A double decomposition-based modelling approach to forecast weekly solar radiation," *Renewable Energy*, vol. 152, pp. 9–22, 2020. https://doi.org/10.1016/j.renene.2020.01.005.

[23] H. Lan, H. Yin, Y.-Y. Hong, S. Wen, C. Y. David, and P. Cheng, "Day-ahead spatio-temporal forecasting of solar irradiation along a navigation route," *Applied Energy*, vol. 211, pp. 15–27, 2018. https://doi.org/10.1016/j.apenergy.2017.11.014.

[24] K. Mohammadi, S. Shamshirband, C. W. Tong, M. Arif, D. Petković, and S. Ch, "A new hybrid support vector machine–wavelet transform approach for estimation of horizontal global solar radiation," *Energy Conversion and Management*, vol. 92, pp. 162–171, 2015. https://doi.org/10.1016/j.enconman.2014.12.050.

[25] S. Monjoly, M. André, R. Calif, and T. Soubdhan, "Hourly forecasting of global solar radiation based on multiscale decomposition methods: A hybrid approach," *Energy*, vol. 119, pp. 288–298, 2017. https://doi.org/10.1016/j.energy.2016.11.061.

[26] X. Qing and Y. Niu, "Hourly day-ahead solar irradiance prediction using weather forecasts by lstm," *Energy*, vol. 148, pp. 461–468, 2018. https://doi.org/10.1016/j.energy.2018.01.177.

[27] H. Acikgoz, "A novel approach based on integration of convolutional neural networks and deep feature selection for short-term solar radiation forecasting," *Applied Energy*, vol. 305, p. 117912, 2022. https://doi.org/10.1016/j.apenergy.2021.117912.

[28] H. Zang, L. Liu, L. Sun, L. Cheng, Z. Wei, and G. Sun, "Short-term global horizontal irradiance forecasting based on a hybrid cnn-lstm model with spatiotemporal correlations," *Renewable Energy*, vol. 160, pp. 26–41, 2020. https://doi.org/10.1016/j.renene.2020.05.150.

[29] P. Kumari and D. Toshniwal, "Long short term memory–convolutional neural network based deep hybrid approach for solar irradiance forecasting," *Applied Energy*, vol. 295, p. 117061, 2021. https://doi.org/10.1016/j.apenergy.2021.117061.

[30] B. Gao, X. Huang, J. Shi, Y. Tai, and J. Zhang, "Hourly forecasting of solar irradiance based on ceemdan and multi-strategy cnn-lstm neural networks," *Renewable Energy*, vol. 162, pp. 1665–1683, 2020. https://doi.org/10.1016/j.renene.2020.09.141.

[31] H. Liu and C. Chen, "Multi-objective data-ensemble wind speed forecasting model with stacked sparse autoencoder and adaptive decomposition-based error correction," *Applied Energy*, vol. 254, p. 113686, 2019. https://doi.org/10.1016/j.apenergy.2019.113686.

[32] P. S. de Mattos Neto, G. D. Cavalcanti, and F. Madeiro, "Nonlinear combination method of forecasters applied to pm time series," *Pattern Recognition Letters*, vol. 95, pp. 65–72, 2017. https://doi.org/10.1016/j.patrec.2017.06.008.

[33] Z. Ma, H. Chen, J. Wang, X. Yang, R. Yan, J. Jia, and W. Xu, "Application of hybrid model based on double decomposition, error correction and deep learning in short-term wind speed prediction," *Energy Conversion and Management*, vol. 205, p. 112345, 2020. https://doi.org/10.1016/j.enconman.2019.112345.

[34] H. Sun *et al.*, "Hybrid model with secondary decomposition, randomforest algorithm, clustering analysis and long short memory network principal computing for short-term wind power forecasting on multiple scales," *Energy*, vol. 221, p. 119848, 2021. https://doi.org/10.1016/j.energy.2021.119848.

[35] Y. Ma, Q. Lv, R. Zhang, Y. Zhang, H. Zhu, and W. Yin, "Short-term photovoltaic power forecasting method based on irradiance correction and error forecasting," *Energy Reports*, vol. 7, pp. 5495–5509, 2021. https://doi.org/10.1016/j.egyr.2021.08.167.

[36] A. Graps, "An introduction to wavelets," *IEEE computational science and engineering*, vol. 2, no. 2, pp. 50–61, 1995. https://doi.org/10.1109/99.388960.

[37] P. Kumari and D. Toshniwal, "Deep learning models for solar irradiance forecasting: A comprehensive review," *Journal of Cleaner Production*, vol. 318, p. 128566, 2021. https://doi.org/10.1016/j.jclepro.2021.128566.

[38] M. E. Torres, M. A. Colominas, G. Schlotthauer, and P. Flandrin, "A complete ensemble empirical mode decomposition with adaptive noise," in *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 4144–4147, IEEE, 2011. https://doi.org/10.1109/ICASSP.2011.5947265.

[39] K. Cho, B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using rnn encoder-decoder for statistical machine translation," *arXiv preprint arXiv:1406.1078*, 2014. https://doi.org/10.48550/arXiv.1406.1078.

[40] L. Xie, Y. Luo, S.-F. Su, and H. Wei, "Graph regularized structured output svm for early expression detection with online extension," *IEEE Transactions on Cybernetics*, pp. 1–13, 2021. https://doi.org/10.1109/TCYB.2021.3108143.

[41] L. Xie, D. Tao, and H. Wei, "Joint structured sparsity regularized multiview dimension reduction for video-based facial expression recognition," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 8, no. 2, pp. 1–21, 2016. https://doi.org/10.1145/2956556.

[42] S. Bai, J. Z. Kolter, and V. Koltun, "An empirical evaluation of generic convolutional and recurrent networks for sequence modeling," *arXiv preprint arXiv:1803.01271*, 2018. https://doi.org/10.48550/arXiv.1803.01271.

[43] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, pp. 1735–1780, 1997. https://doi.org/10.1162/neco.1997.9.8.1735.

[44] H. Sak, A. Senior, and F. Beaufays, "Long short-term memory based recurrent neural network architectures for large vocabulary speech recognition," *arXiv preprint arXiv:1402.1128*, 2014. https://doi.org/10.48550/arXiv.1402.1128.

[45] NREL,Measurement and Instrumentation Data Center. https://midcdmz.nrel.gov/.

[46] DKASC,Alice Spring System. http://dkasolarcentre.com.au/.

[47] M. Kocifaj, "Sky luminance/radiance model with multiple scattering effect," *Solar Energy*, vol. 83, no. 10, pp. 1914–1922, 2009. https://doi.org/10.1016/j.solener.2009.07.004.

[48] F. He, T. Liu, and D. Tao, "Control batch size and learning rate to generalize well: Theoretical and empirical evidence," in *Advances in Neural Information Processing Systems*, vol. 32, 2019. https://dl.acm.org/doi/abs/10.5555/3454287.3454390.

[49] J. Long, D. Yu, G. Wen, L. Li, Z. Wang, and C. P. Chen, "Game-based backstepping design for strict-feedback nonlinear multi-agent systems based on reinforcement learning," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–14, 2022. https://doi.org/10.1109/LSP.2019.2929713.

[50] Y. Li, R. Wang, and Z. Yang, "Optimal scheduling of isolated microgrids using automated reinforcement learning-based multi-period forecasting," *IEEE Transactions on Sustainable Energy*, vol. 13, no. 1, pp. 159–169, 2021. https://doi.org/10.1109/TSTE.2021.3105529.

[51] P. Molchanov, S. Tyree, T. Karras, T. Aila, and J. Kautz, "Pruning convolutional neural networks for resource efficient inference," *arXiv preprint arXiv:1611.06440*, 2016. https://doi.org/10.48550/arXiv.1611.06440.